

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:08:41 ; Search time 6729 Seconds
(without alignments)
12786.661 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggaagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
AK049358
LOCUS

DEFINITION

AK049358 1832 bp mRNA linear HTC 03-APR-2004
Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:C30027E02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC
ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert
sequence.

ACCESSION AK049358

VERSION AK049358.1 GI:26340087

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

10349636

REFERENCE

AUTHORS

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

11042159

REFERENCE

AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

11076861

REFERENCE

AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

TITLE

JOURNAL

REFERENCE

5

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:08:41 ; Search time 6729 Seconds
(without alignments)
12786.661 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_hic:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_est7:
9: gb_gsa1:
10: gb_gsa2:
11: gb_gsa3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799.2	97.8	1832	AK049358	AK049358 Mus muscu
2	1785	97.1	1787	AK086940	AK086940 Mus muscu
3	1780	96.8	1896	AK085617	AK085617 Mus muscu
4	1683	91.5	1684	AK031742	AK031742 Mus muscu
5	1528	83.1	1553	AK005532	AK005532 Mus muscu
6	1388.4	75.5	1621	AK082340	AK082340 Mus muscu
7	1352.8	73.6	1376	AK082749	AK082749 Mus muscu
8	1345.6	73.2	1586	AK010702	AK010702 Mus muscu
9	766	41.7	766	CX166520	CX166520 HES241
10	761.2	41.4	766	BG974240	BG974240 602843972
11	751.4	40.9	926	BQ922065	BQ922065 AGENCOURT
12	743.8	40.4	793	B1104487	B1104487 602890091
13	735.6	40.0	937	CB203728	CB203728 AGENCOURT
14	733.8	39.9	768	BUS22275	BUS22275 AGENCOURT
15	733.6	39.9	892	CF553710	CF553710 AGENCOURT
16	728.8	39.6	903	CB320984	CB320984 AGENCOURT
17	722.2	39.3	985	BF785328	BF785328 602111557
18	713	38.8	737	BG873094	BG873094 602794679
19	711.4	38.7	836	BI078951	BI078951 602873203
20	695.2	37.8	970	CF584966	CF584966 AGENCOURT
21	694.2	37.7	699	CX208584	CX208584 MNS15855
22	691.6	37.6	1330	CR626260	CR626260 full-leng

23	691.6	37.6	1338	4	CR625170	CR625170 full-leng
24	691.6	37.6	1384	4	CR603815	CR603815 full-leng
25	691.6	37.6	1393	4	CR605951	CR605951 full-leng
26	691.6	37.6	1461	4	CR606170	CR606170 full-leng
27	691.6	37.6	1657	4	CR620849	CR620849 full-leng
28	691.6	37.6	1659	4	CR593913	CR593913 full-leng
29	691.6	37.6	1667	4	CR605378	CR605378 full-leng
30	686.2	37.3	1316	4	CR614568	CR614568 full-leng
31	682	37.1	696	6	CD353185	CD353185 UI-M-GLO-
32	678.8	36.9	1140	5	EX375141	EX375141 BX375141
33	678.4	36.9	721	10	AY403772	AY403772 Mus muscu
34	677.8	36.9	942	2	BF385997	BF385997 602046176
35	677	36.8	1024	5	BY710119	BY710119 BY710119
36	672	36.5	991	5	BQ876296	BQ876296 AGENCOURT
37	669.4	36.4	886	5	EX383415	EX383415 BX383415
38	663.8	36.1	1136	3	BM545167	BM545167 AGENCOURT
39	659.2	35.8	1694	4	CR608091	CR608091 full-leng
40	652	35.5	665	6	CF732471	CF732471 UI-M-HAO-
41	646.8	35.2	698	2	B1110829	B1110829 602895470
42	641.8	34.9	779	3	BI688404	BI688404 60316338
43	640.6	34.8	944	5	BUS10421	BUS10421 AGENCOURT
44	638.8	34.7	768	8	DN935262	DN935262 AGENCOURT
45	638.2	34.7	645	2	BB498921	BB498921 BB498921

ALIGNMENTS

RESULT 1	AK049358	AK049358	1832 bp	mrna	linear	HTC 03-APR-2004
LOCUS	AK049358					
DEFINITION	Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone:C330027E02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.					
ACCESSION	AK049358					
VERSION	AK049358.1	GI:26340087				
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1	Carninci, P. and Hayashizaki, Y.				
AUTHORS	High-efficiency full-length cDNA cloning					
TITLE	Meth. Enzymol. 303, 19-44 (1999)					
JOURNAL	10349636					
PUBMED						
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
TITLE	Genome Res. 10 (10), 1617-1630 (2000)					
JOURNAL	11042159					
PUBMED						
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer					
TITLE	Genome Res. 10 (11), 1757-1771 (2000)					
JOURNAL	11076861					
PUBMED						
REFERENCE	4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
AUTHORS	Functional annotation of a full-length mouse cDNA collection					
TITLE	Nature 409, 685-690 (2001)					
JOURNAL						
PUBMED						
REFERENCE	5					

|||||
1229 AAGATGCTCACCAGCGCCCTTCCAGCAGGAAGGACTCCGTGAGGCTCTCAGCGCCCTCAGCGCCAGG 1288
Qy CTTGCCCTCTCTTCCACCAAGCAATCTCTTCTGCTGGTCTCTGATCGGATCGTAAATTCG 1328
Db CTTGCCCTCTCTTCCACCAAGCAATCTCTTCTGCTGGTCTCTGATCGGATCGTAAATTCG 1348
Qy AGAACTTCCAGGACAAACTCGGGTGTGGCACAAGGGGCTGGACCCAGAGCCAGAGCCA 1388
Db AGAACTTCCAGGACAAACTCGGGTGTGGCACAAGGGGCTGGACCCAGAGCCAGAGCCA 1408
Qy CGCCAGAGACTGCGAGAGGGCAGCTGACCTAAACCCCTCGGAAGCCAACTCTCGAGTTC 1448
Db CGCCAGAGACTGCGAGAGGGCAGCTGACCTAAACCCCTCGGAAGCCAACTCTCGAGTTC 1468
Qy CCGTGTCCACCACTCTCTCGAGAGCGCTCATGCTCTGCGCAGCGCTTCTCCAGGGC 1508
Db CCGTGTCCACCACTCTCTCGAGAGCGCTCATGCTCTGCGCAGCGCTTCTCCAGGGC 1528
Qy TACCAGAGTAACACCTTTTGGCCTTCTGGTCTGCTGGTCTCTGAGGCTCCAG 1568
Db TACCAGAGTAACACCTTTTGGCCTTCTGGTCTGCTGGTCTCTGAGGCTCCAG 1588
Qy AGTGTCCCTCATGATCTTTTGGCCTTCTGGTCTGCTGGTCTCTGAGGCTCCAG 1628
Db AGTGTCCCTCATGATCTTTTGGCCTTCTGGTCTGCTGGTCTCTGAGGCTCCAG 1648
Qy CACCATCATTTGGAGGCTTAACTCTGATTTACTAGGAGGCTCTGGAGCGCCGGGTTG 1688
Db CACCATCATTTGGAGGCTTAACTCTGATTTACTAGGAGGCTCTGGAGCGCCGGGTTG 1708
Qy GTTTGGGGTAATCACTCACTGGCTCTCAGCCTTCTAACTGACGCGCCCTTAATACAGTT 1748
Db GTTTGGGGTAATCACTCACTGGCTCTCAGCCTTCTAACTGACGCGCCCTTAATACAGTT 1768
Qy CTTTCTGTTGTGGTACTCCACGCCCCCAGACACACCAATAAATTTTCATGCTG 1808
Db CTTTCTGTTGTGGTACTCCACGCCCCCAGACACACCAATAAATTTTCATGCTG 1828
Qy TTTT 1812
Db TTTT 1832

AK086940 1787 bp mRNA linear HTC 03-APR-2004
Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched library, clone:E030013D02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.
AK086940
AK086940.1 GI:2632264
HTC; CAP trapper.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159
2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1787)
Adachi,J., Aizawa,K., Akimura,T., Atakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saiko,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
Location/Qualifiers
1. .1787
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:E030013D02"
/db_xref="taxon:10090"
/clone="E030013D02"
/tissue_type="lung"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
22. .852
/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828)
putative"
/codon_start=1
/protein_id="BAC39769.1"
/db_xref="GI:2632265"
/translation="MESMSELAPRCLLPILLLLPAPKLGSPSAGAEETDWR LPKCEVCKYAVELKSAFETGKIVTGYGLDGKSGVYKTDLRIEVTET ICKRLLDYSLHKERTGSRNFANGMSEFTETLNLVHKGVKVVMDIPYELMNETSAEVA

ORIGIN									
Query Match 97.1%; Score 1785; DB 4; Length 1787;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
DLKKOCDVLVEEFEEVIEDWYRNHQBEDLTFLCANHLVKGKDTSLAERWSGKGGDI									
ASLGGKSKKRSVGVKSSSGSKQRKELGGLEDANAEBEEGVQRASPLPHSPDEL									
"									
Qy	35	GCTCTGGGTCCGGCTGGGCCATGGAGTCCATGTCGTAGCTCGCGCCCCGCTGCTCTTAT	94						
Db	2	GCTCTGGGTCCGGCTGGGCCATGGAGTCCATGTCGTAGCTCGCGCCCCGCTGCTCTTAT	61						
Qy	95	TTCCCTTTGCTGCTGCTCTCCGCTGCTGCTCTCTTCCTGCGCCGGAAGCTAGGCCGAGTC	154						
Db	62	TTCCCTTTGCTGCTGCTCTCCGCTGCTGCTCTCTTCCTGCGCCGGAAGCTAGGCCGAGTC	121						
Qy	155	CCGCGGGGCTGAGGACACGACTGGGTGGGATTTGCCCGACAAATGCGAAGTGTGCAAGT	214						
Db	122	CCGCGGGGCTGAGGACACGACTGGGTGGGATTTGCCCGACAAATGCGAAGTGTGCAAGT	181						
Qy	215	ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTG	274						
Db	182	ATGTTGCTGTGGAGCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTG	241						
Qy	275	ACACCGGCTATGGCATCTCTGGAACGGGAGGCTCTGGAGTCMAGTACCAAGTCGGACT	334						
Db	242	ACACCGGCTATGGCATCTCTGGAACGGGAGGCTCTGGAGTCMAGTACCAAGTCGGACT	301						
Qy	335	TACGGTTAAATTGAAGTCACTGAGACCAATTTGCAAGAGCTTCTGGACTACAGCCTGCACA	394						
Db	302	TACGGTTAAATTGAAGTCACTGAGACCAATTTGCAAGAGCTTCTGGACTACAGCCTGCACA	361						
Qy	395	AGGAGGAGCTGCGCAGCAACCGGTTTCCCAAGGGTATGTCCGAGACCTTTTGAGACGCTGC	454						
Db	362	AGGAGGAGCTGCGCAGCAACCGGTTTCCCAAGGGTATGTCCGAGACCTTTTGAGACGCTGC	421						
Qy	455	ACAACCTAGTCCACAAAGGGGTCAAGGTGGTGTGATATCCCTATGAGCTGTGGAAAG	514						
Db	422	ACAACCTAGTCCACAAAGGGGTCAAGGTGGTGTGATATCCCTATGAGCTGTGGAAAG	481						
Qy	515	AGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCTGGTGGAAAGTTTG	574						
Db	482	AGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCTGGTGGAAAGTTTG	541						
Qy	575	AAGAGGTGATTGAGGACTGTGTACAGGAAACCAACAGGAGGAGACCTGACTGAATTCCTCT	634						
Db	542	AAGAGGTGATTGAGGACTGTGTACAGGAAACCAACAGGAGGAGACCTGACTGAATTCCTCT	601						
Qy	635	GTGCCAACCAAGTCTGAGGGAAGGACACGAGTTCGCTAGCAGAGGGTGTGCTGGCA	694						
Db	602	GTGCCAACCAAGTCTGAGGGAAGGACACGAGTTCGCTAGCAGAGGGTGTGCTGGCA	661						
Qy	695	AGAAAGGGGACATAGCTCTCCCTGGGAGGAGAAATCCAAAGAAAGCGACGCGAGTCA	754						
Db	662	AGAAAGGGGACATAGCTCTCCCTGGGAGGAGAAATCCAAAGAAAGCGACGCGAGTCA	721						
Qy	755	AGGGCTCCTCAGTGGCAGCAGCAAGCAGAGGAAGGAATGTGGGGGGCTGGGGGAGGATG	814						
Db	722	AGGGCTCCTCAGTGGCAGCAGCAAGCAGAGGAAGGAATGTGGGGGGCTGGGGGAGGATG	781						
Qy	815	CCAAAGCGGAGGAGGAGGAGGTGTGCAAGAGGATGCGCCCTCCACACAGCCCCCTG	874						
Db	782	CCAAAGCGGAGGAGGAGGAGGTGTGCAAGAGGATGCGCCCTCCACACAGCCCCCTG	841						
Qy	875	ATGAGCTGTGAGCCAGCTTAGTGTCTTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	934						
Db	842	ATGAGCTGTGAGCCAGCTTAGTGTCTTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	901						
Qy	935	CACGCACAGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTCCAGCATCAGGTCT	994						
Db	902	CACGCACAGCGCAGCGCAGCTCCAGCAAGGACAGCTGTGTCCAGCATCAGGTCT	961						

Qy	995	CCTCCCTTGCTGTGCCCCCTTCTCTCCCTTGAAACAGCAAGAGTGAAGGATCTGG	1054
Db	962	CCTCCCTTGCTGTGCCCCCTTCTCTCCCTTGAAACAGCAAGAGTGAAGGATCTGG	1021
Qy	1055	GGTCTGGGAGACGGCACCCCAAGGGAAGAGGAGGAGGAGAGGAGGAGCTCTCTTTC	1114
Db	1022	GGTCTGGGAGACGGCACCCCAAGGGAAGAGGAGGAGGAGGAGGAGGAGCTCTCTTTC	1081
Qy	1115	TACACAGTCCCTCTACGAGCTCCGGGTCCACCCAGCATCCCGAGGCTGAGATCCAGGC	1174
Db	1082	TACACAGTCCCTCTACGAGCTCCGGGTCCACCCAGCATCCCGAGGCTGAGATCCAGGC	1141
Qy	1175	TCCTGACATGAAGAGCTGAAGAGCATGAGGCACATAAGATGCTCACAGCGCCCCCTTCAG	1234
Db	1142	TCCTGACATGAAGAGCTGAAGAGCATGAGGCACATAAGATGCTCACAGCGCCCCCTTCAG	1201
Qy	1235	CCAGGAAGGACTCGGTGACGCTCAGCAGCAGGCTTCCTCTTCCACCAAGCATTT	1294
Db	1202	CCAGGAAGGACTCGGTGACGCTCAGCAGCAGGCTTCCTCTTCCACCAAGCATTT	1261
Qy	1295	CTCTTCTGCTGCTCTGTCGGATGTTAAATTCGAGAACTTCCAGGACAAACTCGGGTGT	1354
Db	1262	CTCTTCTGCTGCTCTGTCGGATGTTAAATTCGAGAACTTCCAGGACAAACTCGGGTGT	1321
Qy	1355	GGCACAAAGGGGTGGACGCGCAGAGCCAGACCCAGCAGAGACTGCAGAGAGGCGACCT	1414
Db	1322	GGCACAAAGGGGTGGACGCGCAGAGCCAGACCCAGCAGAGACTGCAGAGAGGCGACCT	1381
Qy	1415	GACCTAAACCCCTGGAAAGCCAAATCTGCAAGTTCCTCCAGTCCACCACTCTCTCTGAGGA	1474
Db	1382	GACCTAAACCCCTGGAAAGCCAAATCTGCAAGTTCCTCCAGTCCACCACTCTCTCTGAGGA	1441
Qy	1475	CGCTCATGCTCTGCCCGCCCTTCTCCAGGGGTACACAGGTAACACCTTTTGGGCTTT	1534
Db	1442	CGCTCATGCTCTGCCCGCCCTTCTCCAGGGGTACACAGGTAACACCTTTTGGGCTTT	1501
Qy	1535	TCGGTTTGGTTCCTGGGTCTCATCAGCTCCAGAGTGTCCCTCATCGATCTTTTGTGC	1594
Db	1502	TCGGTTTGGTTCCTGGGTCTCATCAGCTCCAGAGTGTCCCTCATCGATCTTTTGTGC	1561
Qy	1595	CTTTGTCCCAATCCAGGGGCTGGAAAGGCATCACCATCATTTGGAGGCTTAACTGTCT	1654
Db	1562	CTTTGTCCCAATCCAGGGGCTGGAAAGGCATCACCATCATTTGGAGGCTTAACTGTCT	1621
Qy	1655	AGTTACTAGGAGTGTCTGGAGCGCCCGGGGTGGTTTGGGTAATCACTACCTGGCTCT	1714
Db	1622	AGTTACTAGGAGTGTCTGGAGCGCCCGGGGTGGTTTGGGTAATCACTACCTGGCTCT	1681
Qy	1715	CAGCTTCTAACACTGACGCCCCCTTAAATACAGTTTCTTCTGTGTGTGACTCCACGCCC	1774
Db	1682	CAGCTTCTAACACTGACGCCCCCTTAAATACAGTTTCTTCTGTGTGTGACTCCACGCCC	1741
Qy	1775	CCACACACACACATATAATTTTCGATGCTGTTTCAATACTG	1819
Db	1742	CCACACACACACATATAATTTTCGATGCTGTTTCAATACTG	1786

RESULT 3				
AK085617	1896 bp	mRNA	linear	HTC 03-APR-2004
LOCUS				
DEFINITION	Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630047M13 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.			
ACCESSION	AK085617			
VERSION	AK085617.1	GI:26351704		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1			

Qy 789 GGAACCTGGGGGCTTGGGGAGGATGCAACGCCGAGAGGAGGAGGCTGTGCAGAGGC 848
Db |||||
Qy 816 GGAACCTGGGGGCTTGGGGAGGATGCAACGCCGAGAGGAGGAGGCTGTGCAGAGGC 875
Db |||||
Qy 849 ATGCGCCCTCCACACACAGCCCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGAATC 908
Db |||||
Qy 876 ATGCGCCCTCCACACACAGCCCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGAATC 935
Db |||||
Qy 909 AAGACCCCTGACTTCAGAGCTTGGGACACGCAACGCGCAGGCGAGCCAGCTTCAGCAAG 968
Db |||||
Qy 936 AAGACCCCTGACTTCAGAGCTTGGGACACGCAACGCGCAGGCGAGCTTCAGCAAG 995
Db |||||
Qy 969 GACAGCTGCTGTCCAGCATCAGGTCCTCCTCTGGCTGTGCCCTTTCCTTCCCTTGAA 1028
Db |||||
Qy 996 GACAGCTGCTGTCCAGCATCAGGTCCTCCTCTGGCTGTGCCCTTTCCTTCCCTTGAA 1055
Db |||||
Qy 1029 CAACAGCAAGAGGTGGAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGAAGGGA 1088
Db |||||
Qy 1056 CNAACGCAAGAGGTGGAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGAAGGGA 1115
Db |||||
Qy 1089 GGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCTCCCTCACAGCTCCGGGTCACCC 1148
Db |||||
Qy 1116 GGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCTCCCTCACAGCTCCGGGTCACCC 1175
Db |||||
Qy 1149 CAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGCATGAGGCACAT 1208
Db |||||
Qy 1176 CAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGCATGAGGCACAT 1235
Db |||||
Qy 1209 AAGATGCTCACACGCGCCCTTCAGCGAGGAAGGACTCCGTGACGCTTCAGCAGCCAGG 1268
Db |||||
Qy 1236 AAGATGCTCACACGCGCCCTTCAGCGAGGAAGGACTCCGTGACGCTTCAGCAGCCAGG 1295
Db |||||
Qy 1269 CTGCGCTCTCTTCACCAAGCATCTCTTCTGCTGGTCTTGTTCGATGTTAAATTCG 1328
Db |||||
Qy 1296 CTGCGCTCTCTTCACCAAGCATCTCTTCTGCTGGTCTTGTTCGATGTTAAATTCG 1355
Db |||||
Qy 1329 AGAATTCACAGGACAACTCGGGTGTGCGACAA- GGSGGTGGAAGCCGACAGCCAGAGCC 1387
Db |||||
Qy 1356 AGAATTCACAGGACAACTCGGGTGTGCGACAAAGGGGGTGTGAGCCGACAGCCAGAGCC 1415
Db |||||
Qy 1388 ACGCAGAGACTGCAGAGAGGGCACCTGACCTAACCCCTCCCTGGAAAGCCAATCTGCAGTT 1447
Db |||||
Qy 1416 ACGCAGAGACTGCAGAGAGGGCACCTGACCTAACCCCTCCCTGGAAAGCCAATCTGCAGTT 1475
Db |||||
Qy 1448 CCGTGTGCCACCACTCCTCTGAGGACGCTCATGCTCTGCTCCAGCCCTTCTCCCAAGG 1507
Db |||||
Qy 1476 CCGTGTGCCACCACTCCTCTGAGGACGCTCATGCTCTGCTCCAGCCCTTCTCCCAAGG 1535
Db |||||
Qy 1508 CTACCAAGTAAACACCTTTTGGCTTTTGGCTTTGTTCTGGTCTCTCATCAGCCCTCCA 1567
Db |||||
Qy 1536 CTACCAAGTAAACACCTTTTGGCTTTTGGCTTTGTTCTGGTCTCTCATCAGCCCTCCA 1595
Db |||||
Qy 1568 GAGTGTCCCTCATCGATCTTTTGTGCTTTTGTCTCCCAATCCCAAGGGCTGGAAGGCCA 1627
Db |||||
Qy 1596 GAGTGTCCCTCATCGATCTTTTGTGCTTTTGTCTCCCAATCCCAAGGGCTGGAAGGCCA 1655
Db |||||
Qy 1628 TCACCATCATTTGGAGGCTTAACCTGTGAGTTCATGAGAGGTGCTGGAGGCGCCCGGGGTT 1687
Db |||||
Qy 1656 TCACCATCATTTGGAGGCTTAACCTGTGAGTTCATGAGAGGTGCTGGAGGCGCCCGGGGTT 1715
Db |||||
Qy 1688 GGTTTGGGTTAACTACCTACCTGCTCTCAGGCTTCTAACACTGCAGCCCC--TTAATACA 1745
Db |||||
Qy 1716 GGTTTGGGTTAACTACCTACCTGCTCTCAGGCTTCTAACACTGCAGCCCCCTTTAATACA 1775
Db |||||
Qy 1746 GTTCTCTCTGTGTGGTGACT- CCCAGGCCCCCAACACACACACCAATATAATTTTCGAT 1804
Db |||||
Qy 1776 GTTCTCTCTGTGTGGTGACTCCCCACGCCCCCAACACACACACCAATATAATTTTCGAT 1835
Db |||||
Qy 1805 GCTGTTTCATTAACCTGTAA 1822
Db |||||
Qy 1836 GCTGTTTCATTAACCTGTAA 1853
Db |||||

RESULT 4
AK031742
LOCUS
DEFINITION
Mus musculus 11 days embryo head cDNA, RIKEN full-length enriched library, clone:6230400B2 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.
AK031742
ACCESSION
AK031742.1 GI:26327580
VERSION
HTC; CAP trapper.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1684)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/.

FEATURES

Location/Qualifiers
 1..1621
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:C230039G09"
 /db_xref="taxon:10090"
 /clone="C230039G09"
 /tissue_type="cerebellum"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 230..685

CDS

/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE) (SPR[Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828) putative"
 /codon_start=1
 /protein_id="BAC38471.1"
 /db_xref="GI:26349663"
 /translation="MSTFPELNLVHKVVMVDIYELWNETSAREVADLKKQCDVLVEEFVIEDWIRYHQBEDLFEFLCAHNLKQDTSCLAERWSKGRDIASLGKKSKKRSRGVGVSSSGSKQKELGLGEDANAEVEEGVQKASPLPHSPDDEL"

ORIGIN

Query Match 75.5%; Score 1388.4; DB 4; Length 1621;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1400; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 420 TGCCAAAGGGTATGCGGAGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAA 479
 220 TGCGAAGGGTATGTCGGGACCTTTGAGACGCTGCACAACTAGTCCACAAAGGGGTCAA 279
 480 GGTGGTGATGATATCCCTTATGAGCTGTGGAAAGAGACCTTACAGAGAGGTGGCTGACCT 539
 280 GGTGGTGATGATATCCCTTATGAGCTGTGGAAAGAGACCTTACAGAGAGGTGGCTGACCT 339
 540 CAAGAAGCAGTGTACCTGCTGGTGGAGAGTGTGAGAGGTGATTGAGGACTGGTACAG 599
 340 CAAGAAGCAGTGTACCTGCTGGTGGAGAGTGTGAGAGGTGATTGAGGACTGGTACAG 399
 600 GAACACACAGAGAGAAACCTGACTGAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGAAA 659
 400 GAACACACAGAGAGAAACCTGACTGAATTCCTCTGTGCCAACCAACGCTGCTGAAGGGAAA 459
 660 GGAACAGAGTTGCTTACAGAGCGGTGTGTGGCAAGAGGGGACATAGACCTCCCTGGG 719
 460 GGAACAGAGTTGCTTACAGAGCGGTGTGTGGCAAGAGGGGACATAGACCTCCCTGGG 519
 720 AGGAAGAAATCCAAAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCA 779
 520 AGGAAGAAATCCAAAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCA 579
 780 GCAGAGGAAGAACTGGGGGGCCTTGGGGAGGATGCCAACCGCAGGAGGAGGGGTGT 839

580 GCAGAGGAAGAACTGGGGGGCCTTGGGGAGGATGCCAACCGCAGGAGGAGGGGTGT 639
 840 GCAGAGGCAATCGCCCTCCACACAGCCCCCTTGATGAGCTGTGAGCCAGCTTAGTGT 899
 640 GCAGAGGCAATCGCCCTCCACACAGCCCCCTTGATGAGCTGTGAGCCAGCTTAGTGT 699
 900 CTTTGAATCAAGACCCCTGACTTTTCAGAGCTTTGGGACACGACAGCGAGCGCAGCGAG- 958
 700 CTTTGAATCAAGACCCCTGACTTTTCAGAGCTTTGGGACACGACAGCGAGCGCAGCGAGC 759
 959 CTCAGCAAGACAGCTGCTGTCCAGCATCAGGTCTCTCCCTTGGCTGTGCTCCCTTTCC 1018
 760 CTCAGCAAGACAGCTGCTGTCCAGCATCAGGTCTCTCCCTTGGCTGTGCTCCCTTTCC 819
 1019 TTCCCTTTGAACAACAGCAAGAGGTGGAAGGATCTGGGGTGTCTGGGACAGCGCACCCAAA 1078
 820 TTCCCTTTGAACAACAGCAAGAGGTGGAAGGATCTGGGGTGTCTGGGACAGCGCACCCAAA 879
 1079 GGGAAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCCCTCACAGCTCC 1138
 880 GGGAAAGAGGAGGAGGAGCAGAGGAGCTCTCTTTCTACACAGTCCCCCTCACAGCTCC 939
 1139 GGGTCCACCCAGCATCCCAAGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA 1198
 940 GGGTCCACCCAGCATCCCAAGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA 999
 1199 TGAGGACATCAAGATGCTCACAGCGCCCTTTCAGCCAGGAGGAGTCCGTGAGCGCTC 1258
 1000 TGAGGACATCAAGATGCTCACAGCGCCCTTTCAGCCAGGAGGAGTCCGTGAGCGCTC 1059
 1259 AGCAGCCAGCGCTGCTCTTTTCCACCAAGCATCTCTTTCTGCTGGTCTCTGCGAT 1318
 1060 AGCAGCCAGCGCTGCTCTTTTCCACCAAGCATCTCTTTCTGCTGGTCTCTGCGAT 1119
 1319 GGTAAATTCAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGGCTGGAGCGCCAGA 1378
 1120 GGTAAATTCAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGGCTGGAGCGCCAGA 1179
 1379 GCCAGGCCAGCGCAGAGACTTGACAGAGGAGCCTGACCTTAACCCCTTGGAAAGCCAA 1438
 1180 GCCAGGCCAGCGCAGAGACTTGACAGAGGAGCCTGACCTTAACCCCTTGGAAAGCCAA 1239
 1439 TCTGAGTTCCGCTGCTCCACCCACTCTCTCTGAGAGCGCTCTATGCTCTGCCAGCCCTT 1498
 1240 TCTGAGTTCCGCTGCTCCACCCACTCTCTCTGAGAGCGCTCTATGCTCTGCCAGCCCTT 1299
 1499 CTCCAGGGCTACAGAGTAAAACCTTTTGGCCCTTTGGTTCCTGGGTCTCTCAT 1558
 1300 CTCCAGGGCTACAGAGTAAAACCTTTTGGCCCTTTGGTTCCTGGGTCTCTCAT 1359
 1559 CAGCTCCAGAGTGTCCCTCATCTGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 1618
 1360 CAGCTCCAGAGTGTCCCTCATCTGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 1419
 1619 GGAAGGCCATCACCATCAATTGGAGGCTTAACCTGCTAGTTACTAGGAGGTCTGGAGCG 1678
 1420 GGAAGGCCATCACCATCAATTGGAGGCTTAACCTGCTAGTTACTAGGAGGTCTGGAGCG 1479
 1679 CCAGGGTGTGGTGGGGTAACTCACTGAGGCTCTCAGCCCTTCTAATGAGTGGAGCCCT 1738
 1480 CCAGGGTGTGGTGGGGTAACTCACTGAGGCTCTCAGCCCTTCTAATGAGTGGAGCCCT 1539
 1739 TAATACAGTTCCTTCTGTTGGTGTCTCCACGCCCCCAACACACACACACATAAATTTAT 1798
 1540 TAATACAGTTCCTTCTGTTGGTGTCTCCACGCCCCCAACACACACACACATAAATTTAT 1599
 1799 TTGATGCTGTTTCAATTAATCTGT 1820
 1600 TTGATGCTGTTTCAATTAATCTGT 1621

RESULT 7

AK082749 1376 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus ES cells cDNA, RIKEN full-length enriched library,
 DEFINITION clone:C330002N05 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC
 ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert
 sequence.
 AK082749
 VERSION AK082749.1 GI:26349918
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 1376)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tegan, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 FEATURES
 Location/Qualifiers
 1..1376
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM.DB:C330002N05"
 /db_xref="taxon:10090"
 /clone="C330002N05"
 /cell_type="ES cells"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 <1..521
 CDS
 /note="unnamed protein product; 1600025D17RIK PROTEIN
 (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA
 1600025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID,
 100%length, match=828)
 putative"
 /codon_start=3
 /protein_id="BAC38599.1"
 /db_xref="GI:26349919"
 /translation="KRLLDYSLHKERTGSNRFPAKMSSTFETLHNLVHKGVKVMNDIP
 YELWNETSAEVADLKQCDVLVEFEVIEDWYRNHQBEDLTFELCAHNLVKGKDTSC
 LAERWSGKGDIAISLGSKKSKGRSGVKSSSGSKQRKELGGLGEDANAEBEVGQK
 ASPLPHSPDDEL"
 ORIGIN
 Query Match 73.6%; Score 1352.8; DB 4; Length 1376;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1368; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
 QY 365 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGGAGGACTGGCAGCAACCGGTTTGCCA 424
 DB 1 GCAAGAGGCTTCTGGACTACAGCCTGCAACAGGAGGAGGACTGGCAGCAACCGGTTTGCCA 60
 QY 425 AGGGTATGTCGGAGACCTTTTGAGAGCGTGTCAACACCTAGTCCCAAAAGGGGTCAAGGTGG 484
 DB 61 AGGGTATGTCGGAGACCTTTTGAGAGCGTGTCAACACCTAGTCCCAAAAGGGGTCAAGGTGG 120
 QY 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
 DB 121 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 180
 QY 545 AGCAGTGTGAGCTGCTGCTGGAAGAGTGTGAAGAGGTGATTTGAGGACTGGTACAGAACCC 604
 DB 181 AGCAGTGTGAGCTGCTGCTGGAAGAGTGTGAAGAGGTGATTTGAGGACTGGTACAGAACCC 240
 QY 605 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 664
 DB 241 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 300
 QY 665 CGAGTTGCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCTTCCCTGGGAGGGA 724
 DB 301 CGAGTTGCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCTTCCCTGGGAGGGA 360
 QY 725 AGAATCAAGAAGAGCGGAGGTCAGAGGCTCTCCAGTGGCAGCAGCAGCAGCAGA 784
 DB 361 AGAATCAAGAAGAGCGGAGGTCAGAGGCTCTCCAGTGGCAGCAGCAGCAGCAGA 420
 QY 785 GGAAGGAACCTGGGGGGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGTGTGCAGA 844
 DB 421 GGAAGGAACCTGGGGGGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGTGTGCAGA 480
 QY 845 AGGATGCGCCCTCCCAACAGCCCTCTGATAGCTGTGAGCCAGCTTAGTGTCTTGTG 904
 DB 481 AGGATGCGCCCTCCCAACAGCCCTCTGATAGCTGTGAGCCAGCTTAGTGTCTTGTG 540
 QY 905 AATCAAGACCCCTGACTTACAGGCTTGGGACACCGCAGCGCAGCGCAGCTCCAG 964
 DB 541 AATCAAGACCCCTGACTTACAGGCTTGGGACACCGCAGCGCAGCGCAGCTCCAG 600


```
/clone="2410050022"  
/cell_type="ES cells"  
/clone_lib="RIKEN full-length enriched mouse cDNA library"  
225..1217  
misc_feature  
/note="1600025D17RIK PROTEIN (PUTATIVE RETINOIC  
ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE)  
(SPTR|OSDAU1, evidence: PASTY, 100%ID, 100%length,  
match=828)  
putative"
```

ORIGIN

```
Query Match 73.2%; Score 1345.6; DB 4; Length 1586;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1362; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
  
Qy 420 TGCAAGGGTATGTCGGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAA 479  
Db 215 TGCGAAGGGTATGTCGGAGACCTTTGAGACGCTGCAACACCTAGTCCAAAGGGGTCAA 274  
  
Qy 480 GGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCT 539  
Db 275 GGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCT 334  
  
Qy 540 CAAGAAGCAGTGTGACGCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAG 599  
Db 335 CAAGAAGCAGTGTGACGCTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGACAG 394  
  
Qy 600 GAACACACAGAGGAAGACCTGACTGAATTCCTCTGTGCCAACACACCTGCTGCAAGGGAAA 659  
Db 395 GAACACACAGAGGAAGACCTGACTGAATTCCTCTGTGCCAACACACCTGCTGCAAGGGAAA 454  
  
Qy 660 GGACACAGATTGCTTAGCAGAGCGGTGTGTGGGAAGAGGGGACATAGACCTCCCTGGG 719  
Db 455 GGACACAGATTGCTTAGCAGAGCGGTGTGTGGGAAGAGGGGACATAGACCTCCCTGGG 514  
  
Qy 720 AGGGNAGAAATCCAGAGAGACCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAA 779  
Db 515 AGGGNAGAAATCCAGAGAGACCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAA 574  
  
Qy 780 GCAGAGGAAGAACTGGGGGGGCTGGGGGAGATGCCAACGCCAGGAGGAGGGGTGT 839  
Db 575 GCAGAGGAAGAACTGGGGGGGCTGGGGGAGATGCCAACGCCAGGAGGAGGGGTGT 634  
  
Qy 840 GCAGAAGCATCG-CCCTCCACACACGCCCCCTGATGATGCTGTGAGCCCGAGCTTAGTG 898  
Db 635 GCAGAAGCATCGCCCTCCACACACGCCCCCTGATGATGCTGTGAGCCCGAGCTTAGTG 694  
  
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACGACGACGCGCAGCGCAGCGCAG 958  
Db 695 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACGACGACGCGCAGCGCAGCGCAG 754  
  
Qy 959 CTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCTCTCTGGCTGTGCCCTTTCC 1018  
Db 755 CTCAGCAAGGACAGCTGCTGTCCAGCATCAGGTCTCTCTCTGGCTGTGCCCTTTCC 814  
  
Qy 1019 TTCCCTTGAACACACAGCAAGAGGTGGAAGATCTGGGTGTCTGGGAGACGCAACCCCAA 1078  
Db 815 TTCCCTTGAACACACAGCAAGAGGTGGAAGATCTGGGTGTCTGGGAGACGCAACCCCAA 874  
  
Qy 1079 GGGAGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCCCTCAGAGCTCC 1138  
Db 875 GGGAGAGGAGGAGGAGGAGGAGGAGCTCTCTTCTACAGTCCCCCTCAGAGCTCC 934  
  
Qy 1139 GGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA 1198  
Db 935 GGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCA 994  
  
Qy 1199 TGAGGCACATAGATGCTCACCAGCGCCCTTCCAGCAGGAAGGACTCCGTGACGCTC 1258  
Db 995 TGAGGCACATAGATGCTCACCAGCGCCCTTCCAGCAGGAAGGACTCCGTGACGCTC 1054  
  
Qy 1259 AGCAGCAGGCTGCTCTCTCTCCACAGCATCTCTTCTGCTGCTCTTCTGCTCGAT 1318  
Db 1259 AGCAGCAGGCTGCTCTCTCTCTCCACAGCATCTCTTCTGCTGCTCTTCTGCTCGAT
```

```
Db 1055 AGCAGCAGGCTGCTCTCTCTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCTCGAT 1114  
Qy 1319 GGTAAATTCGAGAACTTCAGGACAAACTCGGGTGTGGCACAAGAGGGGTGAGACCCAGA 1378  
Db 1115 GGTAAATTCGAGAACTTCAGGACAAACTCGGGTGTGGCACAAGAGGGGTGAGACCCAGA 1174  
Qy 1379 GCCAGAGCCAGCCAGAGACTGCGAGAGGACCTGACCTAAACCCCTCGAAAGCCAA 1438  
Db 1175 GCCAGAGCCAGCCAGAGACTGCGAGAGGACCTGACCTAAACCCCTCGAAAGCCAA 1234  
Qy 1439 TCTCAGTTCCCGTGTCCACCCACTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTT 1498  
Db 1235 TCTCAGTTCCCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCCCTT 1294  
Qy 1499 CTCCAGGGCTACAGAGTAAACACCTTTTGGCCCTTTTGGCTTTGGTTGGTCTCTCAT 1558  
Db 1295 CTCCAGGGCTACAGAGTAAACACCTTTTGGCCCTTTTGGCTTTGGTTGGTCTCTCAT 1354  
Qy 1559 CAGCTCCAGAGTGTCCCTCATGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 1618  
Db 1355 CAGCTCCAGAGTGTCCCTCATGATCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 1414  
Qy 1619 GGAAGGCCATCACCACCATTTGAGGCTTAACTGTCTAGTACTAGGAGGTGCTGGAGCG 1678  
Db 1415 GGAAGGCCATCACCACCATTTGAGGCTTAACTGTCTAGTACTAGGAGGTGCTGGAGCG 1474  
Qy 1679 CCCGGGTTGGTTGGGTAATCACTCACTGGCTCTCAAGCTTTCAACACTGCGAGCCCT 1738  
Db 1475 CCCGGGTTGGTTGGGTAATCACTCACTGGCTCTCAAGCTTTCAACACTGCGAGCCCT 1534  
Qy 1739 TAATACAGTCTCTCTGTTGGTGTGCTCCACGCCCCACACACACACCAT 1790  
Db 1535 TAATACAGTCTCTCTGTTGGTGTGCTCCACGCCCCACACACACACCAT 1586
```

RESULT 9

```
CX166520  
LOCUS HESC2_41_D10_g1_A035 NIH_MGC_258 Homo sapiens cDNA clone  
DEFINITION IMAGE:7470381.5', mRNA sequence.  
ACCESSION CX166520  
VERSION CX166520.1 GI:56796600  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 766)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabs-remail.nih.gov  
Tissue Procurement: BresaGen, Inc.  
cDNA Library Preparation: Express Genomics, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM15772 Row: g column: 19  
Seq primer: JENREV (CAGAAACAGCTATGACC)  
High quality sequence stop: 766.  
Location/Qualifiers  
1..766  
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"
```

FEATURES

source

/clone="IMAGE:7470381"
/sex="male"
/tissue_type="human embryonic stem cells differentiated to
an early endodermal cell type"
/cell_type="human embryonic stem cells"
/cell_line="BG01"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="NIH_MGC_258"
/note="vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos and
differentiated to an early endodermal cell type. Cell line
id and NIH Registry designation is BG01. Positive for
GAT44, Mixl1, Mx1, HNF4alpha expression; negative for AFP
expression. Passage number 40. cDNA primed using oligo-dT
primer: 5'-pGACTAGTCTAGATCGGAGCGGCCCT(T)25-3' and
cloned into the EcoRV/NotI sites of pExpress-1. This
primary library is non-normalized (normalized primary
library is NIH MGC 259). It was constructed by Express
Genomics (Frederick, MD). Sequence ends have been trimmed
to exclude vector and regions below Phred quality 16.
Three-prime sequences are presented as their reverse
complement and have been trimmed to exclude polyA. Note:
this is a Mammalian Gene Collection library."

ORIGIN

Query Match 41.7%; Score 766; DB 8; Length 766;
Best Local Similarity 100.0%; Pred. No. 7,7e-179;
Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

36 CTCTGGGTCCGGCTGGGCGATGGAGTCCATCTGTAGCTCGCGCCGCTCCCTCTTATT 95
Db 1 CTCTGGGTCCGGCTGGGCGATGGAGTCCATCTGTAGCTCGCGCCGCTCCCTCTTATT 60

96 TCCTTTGTCTGTCTGCTCCGCTGCTCTCTCTTCCTGCCCGAAGCTAGGCCGAGTCC 155
Db 61 TCCTTTGTCTGTCTGCTCCGCTGCTCTCTCTTCCTGCCCGAAGCTAGGCCGAGTCC 120

156 CGCGGGGCTGAGGAGACCGACTGGGTGCGATTGTCGCCGAAATGCGAGTGTGCAAGTA 215
Db 121 CGCGGGGCTGAGGAGACCGACTGGGTGCGATTGTCGCCGAAATGCGAGTGTGCAAGTA 180

216 TGTTCGTGTGAGCTGAAGTGGCTTTTGAGAAACGGGAAGCAAGGAAGTATTGA 275
Db 181 TGTTCGTGTGAGCTGAAGTGGCTTTTGAGAAACGGGAAGCAAGGAAGTATTGA 240

276 CACCGGCTATGGCATCTCGACGGGAAGGCTCTGGAGTCAAGTACACCAAGTCGACTT 335
Db 241 CACCGGCTATGGCATCTCGACGGGAAGGCTCTGGAGTCAAGTACACCAAGTCGACTT 300

336 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGGGCTTCTGAGTCAAGTACACCAAGTCGACAA 395
Db 301 ACGGTTAATTGAAGTCACTGAGACCATTTGCAAGGGCTTCTGAGTCAAGTACACCAAGTCGACAA 360

396 GGAGAGGACTGCGAGCAACCGGTTTGCCAAAGGATGTGCGAGACCTTTGAGACGCTGCA 455
Db 361 GGAGAGGACTGCGAGCAACCGGTTTGCCAAAGGATGTGCGAGACCTTTGAGACGCTGCA 420

456 CAACCTAGTCCACAAAGGGGTCAAGTGGTGATGATATCCCTATGAGCTGTGGAAACA 515
Db 421 CAACCTAGTCCACAAAGGGGTCAAGTGGTGATGATATCCCTATGAGCTGTGGAAACA 480

516 GACCTCAGCAGAGTGGCTGACCTCAAGAAGAGTGTGACGTGCTGGTGGGAAGATTGA 575
Db 481 GACCTCAGCAGAGTGGCTGACCTCAAGAAGAGTGTGACGTGCTGGTGGGAAGATTGA 540

576 AGAGTGATTGAGGACTGGTACAGGAACCAAGGAGGAACCTGACTGAATTCCTCTG 635
Db 541 AGAGTGATTGAGGACTGGTACAGGAACCAAGGAGGAACCTGACTGAATTCCTCTG 600

636 TGCCAAACCACTGCTGAAGGGAAGGAACACGAGTTGCTTAGCAGAGCGGTGCTTGGCAA 695
Db 601 TGCCAAACCACTGCTGAAGGGAAGGAACACGAGTTGCTTAGCAGAGCGGTGCTTGGCAA 660

696 GAAGGGGACATAGCTCCCTGGGAGGGAAGAAATCAAGAAGAGCGCAGCGAGTCAA 755
Db 661 GAAGGGGGAATAGCTCCCTGGGAGGGAAGAAATCAAGAAGAGCGCAGCGAGTCAA 720

756 GGGCTCTCCAGTGGCAGCAGCAGCAGAGAGAGAACTGGGGGC 801
Db 721 GGGCTCTCCAGTGGCAGCAGCAGCAGAGAGAGAACTGGGGGC 766

RESULT 10
BG974240 766 bp mRNA linear EST 12-JUN-2001
LOCUS BG974240
DEFINITION 602843972F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979944 5',
mRNA sequence.
ACCESSION BG974240
VERSION BG974240.1 GI:14361877
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 766)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10978 row: 0 column: 17
High quality sequence stop: 761.
Location/Qualifiers
1. 766
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4979944"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam4"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996)."

ORIGIN

Query Match 41.4%; Score 761.2; DB 2; Length 766;
Best Local Similarity 99.6%; Pred. No. 1.2e-177;
Matches 763; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

658 AAGGACACGATTCCTAGCAGAGCGGTGGTCTGGCAGAGAGGGGGCATAGCTCCCTG 717
Db 1 AAGGACACGATTCCTAGCAGAGCGGTGGTCTGGCAGAGAGGGGGCATAGCTCCCTG 60

718 GGAGCGAAGAATCAAGAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGC 777
Db 61 GGAGCGAAGAATCAAGAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGC 120

778 AAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAAGCGCCGAGGAGGAGG 837
Db 121 AAGCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAAGCGCCGAGGAGGAGG 180


```
QY 838 GTGCAGAGGCAATCCGCCCTCCACACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGT 897
Db 181 GTGCAGAGGCAATCCGCCCTCCACACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGT 240
QY 898 GTCTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACGACACAGCGCGAGCGCA 957
Db 241 GTCTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACGACACAGCGCGAGCGCA 300
QY 958 GTCTCAGCAAGGACAGCTGCTGTCCAGCATCAAGTCTCTCCCTTGGGCTGTGCCCCCTTC 1017
Db 301 GTCTCAGCAAGGACAGCTGCTGTCCAGCATCAAGTCTCTCCCTTGGGCTGTGCCCCCTTC 360
QY 1018 CTTCCCTTGAACAAACAGCAAGGTTGGAAGATCTGGGGTGTGGGAGACGCGACCCCAA 1077
Db 361 CTTCCCTTGAACAAACAGCAAGGTTGGAAGATCTGGGGTGTGGGAGACGCGACCCCAA 420
QY 1078 AGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1137
Db 421 AGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 1138 CGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGC 1197
Db 481 CGGGTCCACCCAGCATCCAGGCTGAGATCCAGGCTCCTGACATGGAAGCTGAAGAGC 540
QY 1198 ATGAGGCACATAAGATGCTCAACAGCGCCCTTCCAGCAGAGGAGTCCGTTGAGCCT 1257
Db 541 ATGAGGCACATAAGATGCTCAACAGCGCCCTTCCAGCAGAGGAGTCCGTTGAGCCT 600
QY 1258 CAGCAGCAGGCTGCTCTTCTTCCACCAAGCATCTCTCTGCTGCTGCTTGTTCGGA 1317
Db 601 CAGCAGCAGGCTGCTCTTCTTCCACCAAGCATCTCTCTGCTGCTGCTTGTTCGGA 660
QY 1318 TGTTAAATTCGAGAACTTCCAGCAAACTCGGGTGTGGCAAAAGGGCTGGAGCCAG 1377
Db 661 TGTTAAATTCGAGAACTTCCAGCAAACTCGGGTGTGGCAAAAGGGCTGGAGCCAG 720
QY 1378 AGCCAGAGCCAGCGCAGAGACTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1423
Db 721 AGCCAGAGCCAGCGCAGAGACTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 766
```

```
RESULT 11
LOCUS BQ922065 926 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8932400 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6468664
5' mRNA sequence.
ACCESSION BQ922065
VERSION BQ922065.1 GI:22337096
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 926)
NIH-MGC http://mgi.mcgill.ca/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13996 row: i column: 17
High quality sequence stop: 632.
Location/Qualifiers
1..926
/organism="Mus musculus"
```

RESULT 12
B1104487
LOCUS

B1104487 793 bp mRNA linear EST 26-JUN-2001

```
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6468664"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_94"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
```

ORIGIN

```
Query Match 40.9%; Score 751.4; DB 5; Length 926;
Best Local Similarity 99.2%; Pred. No. 3.4e-175;
Matches 755; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 541 AAGAAAGCAGTGTGAGCTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGTGTACAGG 600
Db 1 AAGAAAGCAGTGTGAGCTGCTGTGGAAGAGTTTGAAGAGGTGATTGAGGACTGTGTACAGG 60

QY 601 AACCAACAGGAGGAGACCTGACTGAATTCCTCTGTGCAACCACTGCTGAAGGAAAG 660
Db 61 AACCAACAGGAGGAGACCTGACTGAATTCCTCTGTGCAACCACTGCTGAAGGAAAG 120

QY 661 GACACGATTTGCTAGCAGAGCGGTGTGCTGTGCAAGAGAGGGGACATAGCTCTCCCTGGGA 720
Db 121 GACACGATTTGCTAGCAGAGCGGTGTGCTGTGCAAGAGAGGGGACATAGCTCTCCCTGGGA 180

QY 721 GGGAGAAATCCAAAGAGAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 780
Db 181 GGGAGAAATCCAAAGAGAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 240

QY 781 CAGAGAAAGAACTGGGGGGCTGGGGGAGGATGCAACCGCGAGGAGGAGGAGGAGGAGG 840
Db 241 CAGAGAAAGAACTGGGGGGCTGGGGGAGGATGCAACCGCGAGGAGGAGGAGGAGGAGG 300

QY 841 CAGAAAGCATCGCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 900
Db 301 CAGAAAGCATCGCCCTCCCAACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 360

QY 901 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACGACGACGACGACGACGCT 960
Db 361 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACGACGACGACGACGACGCT 420

QY 961 CCAGCAAGGACAGCTGTGCTCAGCATCAGGTCTCCTCCCTTGGCTGTGCCCCCTTTCCCT 1020
Db 421 CCAGCAAGGACAGCTGTGCTCAGCATCAGGTCTCCTCCCTTGGCTGTGCCCCCTTTCCCT 480

QY 1021 CCCTTGAACAAACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGACGCGCACCCCAAGG 1080
Db 481 CCCTTGAACAAACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGACGCGCACCCCAAGG 540

QY 1081 GAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
Db 541 GAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600

QY 1141 GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCTCCATGATGGAAGCTGAAGAGCATG 1200
Db 601 GGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCTCCATGATGGAAGCTGAAGAGCATG 660

QY 1201 AGGCACATAAGATGCTCAACAGCGCCCTTTCCAGCAGGAGGAGTCCGTGAGGCTCAG 1260
Db 661 AGGCACATAAGATGCTCAACAGCGCCCTTTCCAGCAGGAGGAGTCCGTGAGGCTCAG 720

QY 1261 CAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1301
Db 721 CAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 761
```


1603 CCCAATCCAGGGCTGAAGCCCATACCATCATTTGAGGCTTAACCTGTCAGTTACTA 1662
 541 CCCAATCCAGGGCTGAAGCCCATACCATCATTTGAGGCTTAACCTGTCAGTTACTA 600
 1663 GGAGGTGCTGGAGCGCCCGGGTGTGTTGGGTAATCACTCACTGCTCTCAGCCTTC 1722
 601 NGAGGTGCTGGAGCGCCCGGGTGTGTTGGGTAATCACTCACTGCTCTCAGCCTTC 660
 1723 TAACACTGCAGCCCTTAATACAGTTCCTTCTGTTGGTGAATCCACGCCCCCACCACA 1782
 661 TAACACTGCAGCCCTTAATACAGTTCCTTCTGTTGGTGAATCCACGCCCCCACCACA 720
 1783 CACACCATAAATTT-ATTTCGATGCTGTTTCATCACTGTTAAAAAAA 1828
 721 CACACCATAAATTTAATTTTCATGCTGTTTCAAAAAAAA 767

RESULT 15
 CF553710
 LOCUS
 DEFINITION AGENCOURT_15614073 Organ of Corti Mus musculus cDNA clone
 IMAGE:30519414 5', mRNA sequence.
 CF553710
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.
 1 (bases 1 to 892)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhardt, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Bechara Kachar and Celine Pompeia, NIDCD
 cDNA Library Preparation: Bechara Kachar, Celine Pompeia, NIDCD
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>

Plate: NDAM590 row: 0 column: 07
 High quality sequence stop: 579.
 Location/Qualifiers
 1..892
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30519414"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Organ of Corti"
 /note="Organ: Organ of Corti/Bar; Vector: pBluescriptSK-;
 Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was
 fine dissected from a total of 386 OC as follows: 102
 samples from post-natal (P) day 5; 72 from P6; 60 from P7;
 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24
 from P13. Total RNA was extracted using the micro
 FastTrack kit (catalog # K1593-02; Invitrogen, Carlsbad,
 CA), according to manufacturer's instructions. Reverse
 transcription and library construction were carried out
 with the Uni-Zap XR vector kit (catalog # 237211,
 Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit
 (catalog # 237612), both from Stratagene (La Jolla, CA,
 USA), according to manufacturer's instructions. The
 frequency distribution of the library is as follows: 72%
 of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and
 0.1% 51-150. As to gene function, 45% of genes are present

FEATURES
 Source
 1..892
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30519414"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Organ of Corti"
 /note="Organ: Organ of Corti/Bar; Vector: pBluescriptSK-;
 Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was
 fine dissected from a total of 386 OC as follows: 102
 samples from post-natal (P) day 5; 72 from P6; 60 from P7;
 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24
 from P13. Total RNA was extracted using the micro
 FastTrack kit (catalog # K1593-02; Invitrogen, Carlsbad,
 CA), according to manufacturer's instructions. Reverse
 transcription and library construction were carried out
 with the Uni-Zap XR vector kit (catalog # 237211,
 Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit
 (catalog # 237612), both from Stratagene (La Jolla, CA,
 USA), according to manufacturer's instructions. The
 frequency distribution of the library is as follows: 72%
 of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and
 0.1% 51-150. As to gene function, 45% of genes are present

in GenBank and have know function; 23% have hits in
 GenBank, but do not have assigned function; 12% are
 uncharacterized ESTs and 20% are unidentified. Library
 created in the laboratory of M. Brownstein (NIMH, NIH). A
 complete library description can be found at "

ORIGIN
 Query Match 39.9%; Score 733.6; DB 6; Length 892;
 Best Local Similarity 97.9%; Pred. No. 8.8e-171;
 Matches 753; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
 Qy 47 GCTGGCCATCGAGTCCATGTCGAGTCGCGCCCGCTCTTATTTCTTTGCTGC 106
 Db 7 GGTGGCCATCGAGTCCATGTCGAGTCGCGCCCGCTCTTATTTCTTTGCTGC 66
 Qy 107 TGCTGCTTCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 166
 Db 67 TGCTGCTTCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 126
 Qy 167 AGGAGACCGACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 226
 Db 127 AGGAGACCGACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 186
 Qy 227 AGCTGAAAGTGGCTTTTGAAGGAAACCGGAAAGCAAGGAAGTGAATGACAC 286
 Db 187 AGCTGAAAGTGGCTTTTGAAGGAAACCGGAAAGCAAGGAAGTGAATGACAC 246
 Qy 287 GCATCTCTGGAGCGGAAAGGCTCTGGAGTCAAGTACCAAGTCCGACTTACGGT 346
 Db 247 GCATCTCTGGAGCGGAAAGGCTCTGGAGTCAAGTACCAAGTCCGACTTACGGT 306
 Qy 347 AAGTCACTGAGACCATTTGCAAGAGGCTTTCTGGACTACAGCTTCAAGAGAGG 406
 Db 307 AAGTCACTGAGACCATTTGCAAGAGGCTTTCTGGACTACAGCTTCAAGAGAGG 366
 Qy 407 GCAGCAACCGGTTTGCACAGGTTATGTCGAGACCTTTGAGACCGCTGCACAA 466
 Db 367 GCAGCAACCGGTTTGCACAGGTTATGTCGAGACCTTTGAGACCGCTGCACAA 426
 Qy 467 ACAAGGGGTCNAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 526
 Db 427 ACAAGGGGTCNAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 486
 Qy 527 AGGTGGCTGACCTCAAGAAGCAGTGTGACGCTGCTGGTGGTGGTGGTGGTGGT 586
 Db 487 AGGTGGCTGACCTCAAGAAGCAGTGTGACGCTGCTGGTGGTGGTGGTGGTGG 546
 Qy 587 AGGACTGGTACAGGAACCAACAGGAGGAACTGACTGAATTCCTCTGTGCCAAC 646
 Db 547 AGGACTGGTACAGGAACCAACAGGAGGAACTGACTGAATTCCTCTGTGCCAAC 606
 Qy 647 TGCTGAAGGGAAGGACACGAGTTGCTTACGAGCGGTGGTCTGGCAAGAGGGGACA 706
 Db 607 TGCTGAAGGGAAGGACACGAGTTGCTTACGAGCGGTGGTCTGGCAAGAGGGGACA 666
 Qy 707 TAGC-CTCCCTGGAGGGAAGAAATCCAAAGAAAGCGGACGCGAGTCAAGGGCTCT 765
 Db 667 TAGC-CTCCCTGGAGGGAAGAAATCCAAAGAAAGCGGACGCGAGTCAAGGGCTCT 726
 Qy 766 AGTGGCAGCAGCAAGCAGAGGAAAGAACTGGGGGGCTCTGGGGAGGATG 814
 Db 727 AGTGGCAGCAGCAAGCAGAGGAAAGAACTGGGGGGCTCTGGGGAGGATG 775

Search completed: December 30, 2005, 03:38:20
 Job time : 6737 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:03:20 ; Search time 9079 Seconds
(without alignments)
11513.945 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggaagcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1829.4	99.5	1882	9 BC013549	BC013549 Mus muscu
2	1784.2	97.0	1909	9 AF361644	AF361644 Mus muscu
3	1546.4	84.1	1655	9 AK220209	AK220209 Mus muscu
4	691.6	37.6	1413	8 BC008898	BC008898 Homo sapi
5	691.6	37.6	1413	8 BC008961	BC008961 Homo sapi
6	691.6	37.6	1675	6 AX464002	AX464002 Sequence
7	691.6	37.6	1675	8 AV358960	AV358960 Homo sapi
8	691.6	37.6	1720	8 BC004423	BC004423 Homo sapi
9	502.8	27.3	958	6 CQ716533	CQ716533 Sequence
10	491.2	26.7	962	8 HS080744	HS080744 Homo sapien
11	484	26.3	246404	14 AC131887	AC131887 Rattus no
12	467	25.4	1201	6 AR379595	AR379595 Sequence
13	458.6	24.9	1534	6 AR379729	AR379729 Sequence
14	457.8	24.9	1277	8 BC008133	BC008133 Homo sapi
15	457.8	24.9	1593	8 BC011767	BC011767 Homo sapi
16	446	24.3	1657	6 BD194905	BD194905 86 human
17	446	24.3	1657	6 CQ855233	CQ855233 Sequence
18	416	22.6	222256	14 AC134482	AC134482 Rattus no

19	412.2	22.4	580	6 AX198865	AX198865 Sequence
20	412.2	22.4	580	6 AX209392	AX209392 Sequence
21	359.8	19.6	2623	8 AK124196	AK124196 Homo sapi
22	339.4	18.5	917	5 CR524291	CR524291 Gallus ga
23	331.2	18.0	904	8 BC022093	BC022093 Homo sapi
24	309.2	16.8	1503	5 BC088872	BC088872 Xenopus t
25	300	16.3	758	5 BC083297	BC083297 Danio rer
26	298	16.2	2763	5 BC073680	BC073680 Xenopus l
27	284	15.4	612	8 AF161347	AF161347 Homo sapi
28	266.8	14.5	4469	8 AK090425	AK090425 Homo sapi
29	263.6	14.3	526	6 CQ921127	CQ921127 Sequence
c	186	10.1	89855	14 AC074220	AC074220 Mus muscu
31	172	9.4	1020	9 BC086647	BC086647 Mus muscu
32	172	9.4	1526	6 AX590165	AX590165 Sequence
33	172	9.4	1958	9 BC094672	BC094672 Mus muscu
34	160.2	8.7	744	6 AR142811	AR142811 Sequence
35	160.2	8.7	746	6 BD123856	BD123856 Secretary
36	160.2	8.7	746	6 CS051652	CS051652 Sequence
37	160.2	8.7	746	6 AX136616	AX136616 Sequence
38	160.2	8.7	747	6 BD006701	BD006701 Novel pol
39	160.2	8.7	1413	8 BC019903	BC019903 Homo sapi
40	160.2	8.7	1419	6 BD123666	BD123666 Secretary
41	160.2	8.7	1419	6 CS051459	CS051459 Sequence
42	160.2	8.7	1419	6 AX136423	AX136423 Sequence
43	160.2	8.7	1419	8 AK075537	AK075537 Homo sapi
44	160.2	8.7	1457	6 BD137191	BD137191 Mammalian
45	160.2	8.7	1462	6 BD227243	BD227243 Secreted

ALIGNMENTS

RESULT 1

BC013549

LOCUS

BC013549

DEFINITION

Mus musculus

ACCESSION

BC013549

VERSION

BC013549

KEYWORDS

MGC.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 1882)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buotow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uadin, T.B., Tothiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shervencko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE

human and mouse cdna sequences

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED

12477932

REFERENCE

2 (bases 1 to 1882)

AUTHORS

Director MGC Project.

TITLE

Direct Submission

JOURNAL

Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

BC013549 1882 bp mRNA linear ROD 28-SEP-2004
Mus musculus trinucleotide repeat containing 5, mRNA (cDNA clone MGC:19340 IMAGE:4222133), complete cds.

BC013549

BC013549

MGC.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 1882)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buotow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uadin, T.B., Tothiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shervencko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cdna sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1882)

Director MGC Project.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabos-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRAK Plate: 25 Row: j Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31982623.

FEATURES

```
1. .1882
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/cclone="MGC:19340 IMAGE:422133"
/tissue_type="Kidney, normal. 5 month old male mouse."
/cclone_lib="NCI CGAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
```

gene

```

1. 1892
/ gene="Thrc5"
/ note="synonyms: CAGA, ERDA5"
/ db_xref="GeneID:72039"
/ db_xref="MGI:1919279"
91..921
/ gene="Thrc5"
/ codon_start=1
/ product="trinucleotide repeat containing 5"
/ protein_id="AAH13549.1"

```

CS

```

91.._xref="MGI:1919279"
/genes="Thrc5"
/codon_start=1
/product="trinucleotide repeat containing 5"
/protein_id="AAH13549.1"
/db_xref="GI:15488836"
/db_xref="GeneID:72029"
/db_xref="MGI:1919279"
-translations="MESMESELAPRCLPFLLLLLPLLLLPAKLPGPSAGAETDWR
LPKSCVKYAVELKSAFEETKTYGIDTGYGLDGKSGVKYKSDLRLEVTET
ICRLLDYLHKRETCGNRFKAGMSEFETLHNLVHKGVKVVMDIPELNMETSASVA
DLKKQCDLVAREPEEVIEDWNRHQBEDLTFCLANHVLLKGKDTSCLAERWSSCKGDI
ASLGGKKSKKRGVKGSSSGSSKQKRELGGIGEDANAAEEEGVQKASPLPHSPDDEL

```

ORIGIN

	Query Match	99.5%;	Score 1829.4;	DB 9;	Length 1882;
	Best Local Similarity	99.9%;	Pred. No. 0;		
	Matches 1830;	Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
Qy	9	GGAGGAAGCCGCGCGGGTCCGCTCTGCTCTGGTCCGGCTGGGCCATGGAGTCCATGTC	68		
Db	45	GGAGGAAGCCGCGCGGGTCCGCTCTGCTCTGGTCCGGCTGGGCCATGGAGTCCATGTC	104		
Qy	69	TGAGCTCGCGCCCGCGTGCCTCTTATTTCTTTGTGCTGCTGCTGCTCCGCTGCTGCTCCT	128		
Db	105	TGAGCTCGCGCCCGCGTGCCTCTTATTTCTTTGTGCTGCTGCTGCTCCGCTGCTGCTCCT	164		
Qy	129	TCCTGCCGCCGAAGCTAGGCGCCGAGTCCGCGCGGGGCTGAGCAGACCGACTGGCGTGCAGATT	188		
Db	165	TCCTGCCGCCGAAGTAGGCGCCGAGTCCGCGCGGGGCTGAGCAGACCGACTGGCGTGCAGATT	224		
Qy	189	GCCACGAAATGGAAGTGTGCAAGTATGTTCTGTGGAGCTGAAGTCGGCTTTTGAGGA	248		
Db	225	GCCACGAAATGGAAGTGTGCAAGTATGTTCTGTGGAGCTGAAGTCGGCTTTTGAGGA	284		
Qy	249	AACGGGAAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCGTCGGAGCGGAAGGGCTC	308		

QY 517 ACCTCAGCAGAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGGTGGAGAGTTTGAA 576
 Db 361 ACCTCAGCAGAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGGTGGAGAGTTTGAA 420
 QY 577 GAGGTGATTGAGGACTGGTACAGGAACACACAGGAGGAGACCTGACCTGAATTCCTCTGT 636
 Db 421 GAGGTGATTGAGGACTGGTACAGGAACACACAGGAGGAGACCTGACCTGAATTCCTCTGT 480
 QY 637 GCCAACACGCTGCTGAAGGGAAGGACACGAGTTGCTTACGAGAGCGGTGGTCTGGCAAG 696
 Db 481 GCCAACACGCTGCTGAAGGGAAGGACACGAGTTGCTTACGAGAGCGGTGGTCTGGCAAG 540
 QY 697 AAGGGGACATAGCTCTCCCTGGGAGGGAGGAAATCAAGAGAGCGGAGCGGAGTCAAG 756
 Db 541 AAGGGGACATAGCTCTCCCTGGGAGGGAGGAAATCAAGAGAGCGGAGCGGAGTCAAG 600
 QY 757 GGCTCTCCAGTGGCAGCAGCAGCAGAGGAGGAACTGGGGGGCTGGGGGAGATGCC 816
 Db 601 GGCTCTCCAGTGGCAGCAGCAGCAGAGGAGGAACTGGGGGGCTGGGGGAGATGCC 660
 QY 817 AACGCC---GAGGAGGAGGAGGTGTGCAAGAGGATCGCCCTCCACACAGACCCCTCT 873
 Db 661 AACGCCGAGGAGGAGGAGGTGTGCAAGAGGATCGCCCTCCACACAGACCCCTCT 720
 QY 874 GATGAGTGTGAGCCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGG 933
 Db 721 GATGAGTGTGAGCCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGG 780
 QY 934 ACAGCAGCGCAGCGCAGCGCAGCTCCAGCAGGAGCAGTGTGCTCCAGCATCAGTGC 993
 Db 781 ACAGCGA-----CAGCAGCGCAGCTCCAGCAGGAGCAGTGTGCTCCAGCATCAGTGC 835
 QY 994 TCCTCCCTTGGTGTGCCCCCTTCTCCCTTGAACAAACAGCAAGAGTGAAGATCTG 1053
 Db 836 TCCTCCCTTGGTGTGCCCCCTTCTCCCTTGAACAAACAGCAAGAGTGAAGATCTG 895
 QY 1054 GGGTCTGGGAGACGGCACCCCAAGGGAAGGAGGAGGAGCAAGCGCAGCTCTCTTT 1113
 Db 896 GGGTCTGGGAGACGGCACCCCAAGGGAAGGAGGAGGAGCAAGCGCAGCTCTCTTT 955
 QY 1114 CTACACAGTCCCTCAGCAGTCCGGGTGACCCAGCATCCCGAGCTGATCCAGG 1173
 Db 956 CTACACAGTCCCTCAGCAGTCCGGGTGACCCAGCATCCCGAGCTGATCCAGG 1015
 QY 1174 CTCCTGACATGGAAGCTGAAGAGCATGAGGCACATAAGATGCTACACAGCGCCCTTCA 1233
 Db 1016 CTCCTGACATGGAAGCTGAAGAGCATGAGGCACATAAGATGCTACACAGCGCCCTTCA 1075
 QY 1234 GCCAGGAAGGACTCGCTGAGCCTCAGCAGCGCAGCGCTTCTTCTTCCACCAAGCAT 1293
 Db 1076 GCCAGGAAGGACTCGCTGAGCCTCAGCAGCGCAGCGCTTCTTCTTCCACCAAGCAT 1135
 QY 1294 TCTCTTCTGCTGCTTGTGGATGGTAAATTCAGAACTTCCAGGACAACTCGGGTG 1353
 Db 1136 TCTCTTCTGCTGCTTGTGGATGGTAAATTCAGAACTTCCAGGACAACTCGGGTG 1195
 QY 1354 TGGCACAAGGGGCTGGAGCGCAGAGCCAGAGCCAGAGCTGAGAGGAGGACCC 1413
 Db 1196 TGGCACAAGGGGCTGGAGCGCAGAGCCAGAGCCAGAGCTGAGAGGAGGACCC 1255
 QY 1414 TGACCTAACCCCTCGGAAGCCATCTGCGAGTTCCTCGTGTCCAGCCACTCTCTCTGAGG 1473
 Db 1256 TGACCTAACCCCTCGGAAGCCATCTGCGTGTTCCTCGTGTCTCTCTCTGAGG 1315
 QY 1474 AGGCTCATGCTTGGCCAGCGCTTCTCCAGGGGTACAGAGTAAACACCTTTTGGCT 1533
 Db 1316 AGGCTCATGCTTGGCCAGCGCTTCTCCAGGGGTACAGAGTAAACACCTTTTGGCT 1375
 QY 1534 TTCGGTTTGGTTCCTGGGTCTCATCAGCTCCAGAGTGCCTTCATCGATCTTTTTTG 1593
 Db 1376 TTCGGTTTGGTTCCTGGGTCTCATCAGCTCCAGAGTGCCTTCATCGATCTTTTTTG 1435

QY 1594 CCTTTGTCCTCCCAATCCAGGGGCTGGAGGCAATCACCATCATTTGGAGGCTTAACCTGT 1653
 Db 1436 CCTTTGTCCTCCCAATCCAGGGGCTGGAGGCAATCACCATCATTTGGAGGCTTAACCTGT 1495
 QY 1654 CAGTTACTAGAGGTGCTGGGAGCGCCCGGGTTCGTTTGGGGTAATCACTCACTGGCTC 1713
 Db 1496 CAGTTACTAGAGGTGCTGGGAGCGCCCGGGTTCGTTTGGGGTAATCACTCACTGG--- 1552
 QY 1714 TCAGCCTTTAACTGCACTGCAAGCCCTTAATAAGATTCCTTCTGTTGTTGAGTCC--- 1768
 Db 1553 -----TTTCTAACTGCAAGCCCTTAATAAGATTCCTTCTGTTGTTGAGTCCCCCAAC 1607
 QY 1769 -CAGCCCGCCACACACACACCAATAAATTA-TTTCGATGCTGTTTCAT 1814
 Db 1608 GCCCCCAACCCACACACACCAATAAATTA-TTTCGTTGCTGTTTCAT 1655

RESULT 4

BC008898 1413 bp mRNA linear PRI 29-JUN-2004
 Homo sapiens trinucleotide repeat containing 5, mRNA (cDNA clone
 MGC:1220 IMAGE:2959532), complete cds.

ACCESSION

BC008898

VERSION

BC008898.2 GI:33870484

KEYWORDS

MGC.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 1413)

AUTHORS

Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL

PUBMED

REFERENCE

2 (bases 1 to 1413)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (29-MAY-2001)

National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK

On Aug 19, 2003 this sequence version replaced gi:14286203.

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org

contact: amadan@systemsbio.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kerteman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Qy	5	CGAGGGAGGAAGCGCCCGCGGTCCGCTCTGCTCTGGGTCCGGCTCGGGCCATGGAGTCCA	64
Db	19	CGCGGGAGGAGGAACCCGCCCGTCTTTTAGGGTCCGGGCCCGGGCCATGGATTCA	78
Qy	65	TGTCGTAGCTCGCGCCCGCTGCCCTCTATTTCCTTTTGCTGCTGCTCTCCGCTGCTGC	124
Db	79	TGCTGAGCCCGCGTCCCGCTGCTCTCTGCTTTCTTCCCTTGCTGCTGCTGCTGCTGC	138
Qy	125	TCCTTCCTGCCCGAAGCTAGCCCGGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC	184
Db	139	TGCTGCCGGGCCCGGAGCTGGGCCCGGACGAGGCCGAGCTGAGGAGAACGACTGGGTTC	198
Qy	185	GATTGCCCGACCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTCAAGTCGGCTTTTG	244
Db	199	GCCTGCCCGACCAATGCGAAGTGTGTAATAATGTTGCTGTGGAGCTGAGTCAAGTCAG	258
Qy	245	AGGAAACGGGAAGACCAAGAAAGTATGTGACACCGGCTATGGCATCTCTGACGGGAAG	304
Db	259	AGGAAACCGGCAAGACCAAGAGAGTATGTGCAACGGGCTATGGCATCTCTGACCAAGG	318
Qy	305	GCCTCTGAGTCAAGTACACCAAGTCGACATTCACGGTTAAATTGAAGTCACCTGAGACCA	364
Db	319	CCTCTGAGTCAATATACCAAGTCGACCTTCGGTTAATCGAAGTCACTTGAGNACCA	378
Qy	365	GCAAGAGGCTCTGGACTACAGCTGCAACGAGGAGGACTGGCGAGCAACCGGTTTCCCA	424
Db	379	GCAAGAGGCTCTTGCAATTATAGCTGCAACGAGGAGGACCGGCGAGCAATCGATTTC	438
Qy	425	AGGATGTCGGAGACCTTTGAGACGCTGCACACCTAGTCCACAAAGGGGTCAAGTGG	484
Db	439	AGGCAATGTCAAGACCTTTGAGACATTCACAACTTGGTATACAAAGGGGTCAAGTGG	498
Qy	485	TGATGGATATCCCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGGTGGCTGACCTCA	544
Db	499	TGATGGACATCCCCCTATGAGCTGTGGAAACGAGATTTCTGACAGGTTGGCTGACCTCA	558

[illegible]

Matches	791;	Conservative	0;	Mismatches	139;	Indels	6;	Gaps	1;
Qy	5	CGAGGGAGGAGCGCGCGCGGTCTCTCTCTCTGGTCCGGTGGGCCATGAGTCCA	64						
Db	287	CGCGGGAGGAGGAACCGCGCGGTCTTTAGGGTCCGGGCCCGCGGCCATGATCAA	346						
Qy	65	TGCTGTGAGCTCGCGCGCGCGTCTCTATTTCTTCTGCTCTGCTCTCGCTGTGTC	124						
Db	347	TGCTGTGAGCTCGCGCGCGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	406						
Qy	125	TCCTTCTGTCGCCGAGCTAGCGCCGAGTCCCGCGGGGCTGAGAGACCGACTGGGTGC	184						
Db	407	TGCTGCGCGCCCGAGCTGGGCGCGAGCCAGCGGAGCTGAGAGAAACGACTGGGTTC	466						
Qy	185	GATTGCCAGCAATGCGAGCTGCGAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG	244						
Db	467	GCCTGCCAGCAATGCGAGTGTAAATATGTTGCTGTGAGCTGAAGTCAGCCCTTG	526						
Qy	245	AGGAAACGGGAAAGACCAAGGAAGTGTATGACACCGGCTATGGCATCTGGACCGGAAGG	304						
Db	527	AGGAAACCGGCAAGACCAAGAGGTGATTGGCACCGGCTATGGCATCTGGACCAAGG	586						
Qy	305	GCTCTGGAGTCAAGTACACCAAGTCGGACTTACCGTTAATTGAAGTCACTGAGACCATTT	364						
Db	587	CCTCTGGAGTCAATACACCAAGTCGGACTTTCGGTTAATCGAAGTCACTGAGACCATTT	646						
Qy	365	GCAAGAGCTTCTGGACTACAGCTGCACAAGAGAGGACTGGCAGCAACCGGTTTGCCA	424						
Db	647	GCAAGAGCTCTGTGATTATACCTTGCAACAAGGAGAGACCGGCAAGTCAATCGATTGCCA	706						
Qy	425	AGGGTATGTCGGAGACCTTTGAGAGCGTGCACAACCTAGTCCACAAGGGGTCAAGGTGG	484						
Db	707	AGGGCATGTCAAGACCTTTGAGACATTACACAACCTGGTATACACAAGGGGTCAAGGTGG	766						
Qy	485	TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGTGGCTGACCTCAAGA	544						
Db	767	TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826						
Qy	545	AGCAGTGTGACGTGCTGTGGAAGATTTGGAAGGTGATTGAGGACTGGTACAGGACCC	604						
Db	827	AGCAGTGTGATGTGCTGTGGAAGATTTGAGGAGGTGATCGAGGACTGGTACAGGACCC	886						
Qy	605	ACCAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACAGCTGCTGAAGGGAAGGACA	664						
Db	887	ACCAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACAGCTGCTGAAGGGAAGGACA	946						
Qy	665	CGAGTTGCTTAGCAGAGCGGTGTTGTCGAAGAAGGGGACATAGCCTCCCTGGGAGGGA	724						
Db	947	CCAGTTGCTTGCAGAGCAGTGGTCCGCAAGAGGGAGACACAGCTGCCCTGGGAGGGA	1006						
Qy	725	AGAAATCCAAGAAGCGCAGCGAGTCAAG-----GGCTCTCTCAGTGGCAGCAGCA	778						
Db	1007	AGAAATCCAAGAAGAGCAGCAGCGGCCCAAGGCAGCAGCGGCGGAGGAGTAGCAGCAGCA	1066						
Qy	779	AGCAGGAGGAGACTGGGGGCGCTGGGGGAGGATGCCAACCGCAGGAGGAGGAGGGTG	838						
Db	1067	AACAAGGAGGAGCTGGGTGGCTTTGAGGAGACCCCGAGGAGGATGAGGCA	1126						
Qy	839	TGCAAGGCGCATCGCCCTCCACACACAGCCCGCTGTGATGAGTGTGAGCCCGAGCTTAGTG	898						
Db	1127	TCCAGAGGCGATCCCTCTCACACACAGCCCGCTGTGATGAGTCTGTAGCCCGAGCAT	1186						
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA	934						
Db	1187	CCTCTGCTCAGACCCCTGATTTTGAAGCTGAGGA	1222						
RESULT	7								
AY358960									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									

KEYWORDS	FLI CDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
	Hominidae; Homo.
REFERENCE	1 (bases 1 to 1675)
AUTHORS	Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J.,
	Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B.,
	Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E.,
	Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S.,
	Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C.,
	Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V.,
	Stinson,J., Vagts,A., Vanden,R., Watanabe,C., Wealand,D., Woods,K.,
	Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z.,
	Goddard,A., Wood,W.I. and Godowski,P.
TITLE	The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
	Effort to Identify Novel Human Secreted and Transmembrane Proteins:
	A Bioinformatics Assessment
	Genome Res. 13 (10), 2265-2270 (2003)
REFERENCE	2 (bases 1 to 1675)
AUTHORS	Clark,H.F.
TITLE	Direct Submission
JOURNAL	Submitted (01-AUG-2003) Department of Bioinformatics, Genentech,
FEATURES	Inc., 1 DNA Way, South San Francisco, CA 94080, USA
source	Location/Qualifiers
gene	1..1675
CDS	/locus_tag="UNQ1934"
	337..1173
	/locus_tag="UNQ1934"
	/note="PRO4409"
	/codon_start=1
	/product="CTG4A"
	/protein_id="AA089319.1"
	/db_xref="GI:37183038"
	/translation="MDSMPFASRCLLLPLLLLLLLLPAPELGPQAGABENDWVR
	LPSCRYCKYAVELKSAFETGKTEVIGTYGILDOKASGVKTKDRLIEVET
	ICKRLDYSILHKERTGSRNFAKMSETFETLHNLVHKGVKVMIDPYELWNETSAEVA
	DLKKCDVLVEEREVEIWDYRHOEDLTFLCANHLVKDTSCLAEOWSGKKGDT
	ALGKGKSKKSSRAKAGGRSSSKQKELGLEGDPSPDEDSGIQKASPLTHSPPD
ORIGIN	EL"
Query Match	37.6%; Score 691.6; DB 8; Length 1675;
Best Local Similarity	84.5%; Pred. No. 1.9e-158;
Matches 791; Conservative	0; Mismatches 139; Indels 6; Gaps 1;
Qy	5 CGAGGGAGGAGCGCGCGCGGTCTCTCTCTCTGGTCCGGTGGGCCATGAGTCCA 64
Db	287 CGCGGGAGGAGGAACCGCGCGGTCTTTAGGGTCCGGGCCCGCGGCCATGATCAA 346
Qy	65 TGCTGTGAGCTCGCGCGCGTCTCTATTTCTTCTGCTCTGCTCTCGCTGTGTC 124
Db	347 TGCTGTGAGCTCGCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 406
Qy	125 TCCTTCTGTCGCCGAGCTAGCGCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184
Db	407 TGCTGTGCGCGCCCGAGCTGGGCGCGAGCCAGCGGAGCTGAGGAGAACGACTGGGTTC 466
Qy	185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db	467 GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCCTTTG 526
Qy	245 AGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTGGACCGGAAGG 304
Db	527 AGGAAACCGGCAAGACCAAGAGGTGATTGGCACCGGCTATGGCATCTGGACCAAGG 586

Db 338 CGCGGAGGAGAAACCGCCGCTCTTTAGGTCGCGGCCCGCGGCATGATTCAA 397
QY 65 TGTCTGAGCTCGCGCCCGCTCTTTATTTCTTTCTGCTGCTTCCGCTGCTGC 124
Db 398 TGCCTGAGCCCGCGCTGCTTCTGCTTCTTCCCTTCTGCTGCTGCTGCTGC 457
QY 125 TCCCTTCTGCCCCAAGCTAGCGCCGAGTCCCGCGCGGCTGAGGAGACCGACTGGGTGC 184
Db 458 TGCTCCCGCCCCGAGCTGGGCCGAGCCAGCGGAGCTGAGGAGAACGACTGGGTTC 517
QY 185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGAGTGAAGTCGGCTTTTG 244
Db 518 GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGTGAAGTCAGCCTTTG 577
QY 245 AGGAACCGGAAAGAACCAAGGATGATTGACACCGGCTATGCGATCTGACCGGAGG 304
Db 578 AGGAACCGGCAAGAACCAAGGAGTGATTGGACCGGCTATGGCATCTGGACAGAAAG 637
QY 305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTT 364
Db 638 CTTCTGGAGTCAATATACCAAGTCGGACTTTCGGTTAATCGAAGTCACTGAGACCAATTT 697
QY 365 GCAAGAGCTTCTGAGCTACAGCTGCAAGAGGAGGAGTGGCAGCAACCGGTTTGCCA 424
Db 698 GCAAGAGCTCTCTGATTATAGCCTGCAAGAGGAGGAGCCGGCAGCAATCGATTGCGCA 757
QY 425 AGGATATGTCGAGACCTTTGAGAGCTGCAACACCTAGTCCACAAAGGGGTCAAGGTGG 484
Db 758 AGGCGATGTCAGAGACCTTTGAGACATTTACACACCTGGTATCACAAAGGGGTCAAGGTGG 817
QY 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 818 TGATGGACATCCCTATGAGCTGTGGAACGAGACTTCTGAGAGGTGCTGACCTCAAGA 877
QY 545 AGCAGTGTGCTGCTGTGGAAGTTTGAAGAGTGAATGAGAGCTGGTACAGGAACC 604
Db 878 AGCAGTGTGATGCTGTGGAAGTTTGAAGAGTGTGAGGAGTGTATCGAGACTGGTACAGGAACC 937
QY 605 ACCAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAAGGGAAGGACA 664
Db 938 ACCAGGAGGAAGACCTGACTGAATTCCTCTGCGCCAACCACTGCTGAAGGGAAGGACA 997
QY 665 CGAGTTGCTAGCAGAGCGGTGCTGGAAGAGGGGACATAGCTCTCCCTGGGAGGGA 724
Db 998 CAGTTGCTGCGAGAGCAGTGTGTCGCAAGAGGGAGACACAGCTGCTGGGAGGGA 1057
QY 725 AGAAATCAAGAAGAGCGAGCGAGTCAAG-----GGTCTCTCAGTGGCAGCAGCA 778
Db 1058 AGAAGTCCAAGAAGAGAGCAGCAGGCGCAAGGCAGCGGCGGAGGATGACAGCAGCA 1117
QY 779 AGCAGAGGAAGAACTGGGGGCTTGGGGGAGGATGCCAACGCCGAGGAGGAGGGGTG 838
Db 1118 AACAAAGGAAGAGCTGGGTGGCTTGGGAGAGACCCAGCGCCCGGAGGAGATGAGGGA 1177
QY 839 TGCAGAAGGCATCGCCCTCCACACAGCCCCCTGATGAGTGTGAGCCCGAGCTTAGTG 898
Db 1178 TCCAGAAGGCATCCCTCTCAACACAGCCCCCTGATGAGCTCTGAGCCCCACCAGCAT 1237
QY 899 TCCTTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 1238 CTTCTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 1273

RESULT 9
LOCUS CQ716533 958 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 2467 from Patent WO2068579.
ACCESSION CQ716533
VERSION CQ716533.1 GI:42277390
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
Venter, C.J., Adams M.C., Li, P.W. and Myers E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
Patent: WO 02068579-A 2467 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source Location/Qualifiers
1..958
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 27.3%; Score 502.8; DB 6; Length 958;
Best Local Similarity 87.7%; Pred. No. 4e-112;
Matches 549; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 5 CGAGGAGGAGAACCGCGCGGTCCTCTCTGCGTCCGGCTGGGCCCATGGAAGTCCA 64
Db 333 CGCGGAGGAGAAACCGCCCGTCTTTAGGGTCGCGCCCGCGGCCCATGGAATCAA 392
QY 65 TGTCTGAGCTCGCGCCCGCTCTCTTTATTTCTTTCTGCTGCTGCTTCCGCTGCTGC 124
Db 393 TGCCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTTCTGCTGCTGCTGCTGC 452
QY 125 TCCCTTCTGCCCCAAGACTAGGCCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184
Db 453 TGTCTCGCGGCCCGGAGCTGGGCCCGAGCCAGCGGAGCTGAGGAGAACGACTGGGTTC 512
QY 185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db 513 GCCTGCCAGCAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 572
QY 245 AGGAACCGGAAAGAACCAAGGAGTGAATTGACACCGGCTATGGCATCTTGGACGGAAGG 304
Db 573 AGGAACCGGCAAGAACCAAGGAGGTGATTGGCACGGGCTATGGCATCTGGACCAAGAAGG 632
QY 305 GCTCTGAGTCAAGTACACCAAGTCGGACTTACGGTTAATTGAAGTCACTGAGACCAATTT 364
Db 633 CTTCTGGAGTCAAAATACCAAGTCGGACTTTCGGTTAATCGAAGTCACTGAGACCAATTT 692
QY 365 GCAGAGGCTTCTGAGCTACAGCTGCAACAGGAGAGGAGCTGGCAGCAACCGGTTTGCCA 424
Db 693 GCAAGAGCTCTCTGATTTATAGCTTGCACAAAGGAGAGGACCGCGCAGCAATCGATTGCGA 752
QY 425 AGGCTATGTCGGAGACCTTTGAGACGCTGCAACAACCTAGTCCCAAAAGGGGTCAAGGTGG 484
Db 753 AGGCGATGTCAGAGACCTTTGAGACATTTACACAACCTGGTACACAAGGGGTCAAGGTGG 812
QY 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 813 TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGCGAGAGGTGGCTGACCTCAAGA 872
QY 545 AGCAGTGTGAGCTGCTGTTGGAAGAGTTTGAAGAGTGAATGAGGAGCTGATGAGGAGGAGT 604
Db 873 AGCAGTGTGATGCTGTTGGAAGAGTTTGAAGAGGTGATCGAGACTGGTACAGGAACC 932
QY 605 ACCAGGAGGAAGACCTGACTGAATTC 630
Db 933 ACCAGGAGGAAGACCTGACTGAATTC 958
RESULT 10
LOCUS HSU80744
DEFINITION Homo sapiens CTG4a mRNA, complete cds.
ACCESSION U80744
VERSION U80744.1 GI:2565062
KEYWORDS
SOURCE Homo sapiens (human)


```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S.,
Brechel, T.S., Stine, O.C., Callahan, C., McInnis, M.G. and Ross, C.A.
TITLE cDNAs with long CAG trinucleotide repeats from human brain
JOURNAL Hum. Genet. 100 (1), 114-122 (1997)
PUBMED 9225980
REFERENCE
AUTHORS Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S.,
Brechel, T.S., Stine, O.C., Callahan, C., McInnis, M.G. and Ross, C.A.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES
source
1. .962
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="cerebral cortex"
1. .962
/gene="CTG4a"
388. .819
/gene="CTG4a"
/feature="poly-leucine rich"
/codon_start=1
/product="CTG4a"
/protein_id="AAB91442.1"
/db_xref="GI:2565063"
/translation="MDSMPSPASRCLLLPILLLLPAPLPGSPQAGAEENDWVR
LPSECKVYVAVELKSAFETGKTVIGTGYGLDOKASGVKTKYKSDLRLEIETET
ICKLLLIACTRRGPAIDLPACQRLRHVTWTYKGRW"
repeat_region 439..462
/rpt_type=tandem
/rpt_unit="ctg"

ORIGIN
Query Match 26.7%; Score 491.2; DB 8; Length 962;
Best Local Similarity 87.3%; Pred No. 2.8e-109;
Matches 549; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 2 GCACGAGGAGGAGGAGCGCGCGGCTCTGCTCTGGGTCCGGCTGGGCCATGGAGT 61
DB 335 GCGCGCGGAGGAGGAGCGCGCGTCTTAGGGTCCGGCGCGCGCGCATGGATT 394

QY 62 CCAATGCTGAGTCCGCGCGCGCTGCTCTTATTTCTTGTGCTGCTGCTGCTGCTGC 121
DB 395 CAATGCTGAGCGCGCGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 454

QY 122 TGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 181
DB 455 TGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 514

QY 182 TGCGATTCTCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCCGCTT 241
DB 515 TTGCGCTCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCCGCT 574

QY 242 TTGAGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGGACGGGA 301
DB 575 TTGAGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGGACGGGA 634

QY 302 AGGGCTCTGGAGTCAAGTACCAAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCC 361
DB 635 AGGGCTCTGGAGTCAAGTACCAAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCC 694

QY 362 TTGCAAGAGGCTCTCTGAGCTACAGCTGCAAGGAGGAGTCCGAGTCCGAGTCCGAGT 421
DB 695 TTGCAAGAGGCTCTCTGAGCTACAGCTGCAAGGAGGAGTCCGAGTCCGAGTCCGAGT 753

QY 422 CCAAGGGTATGTGCGGAGACCTTTGAGACGCTGCACACCTAGTCCACAAAGGGGTCAAGG 481

```

```

Db 754 CCAAGGGATGTGAGAGACCTTTGAGACATTTACACAACTGTGACAAAGGGGTCAAGG 813
QY 482 TGGTGATGATATCCCTTATGAGTGTGGAACGAGACCTTCAGCAGAGTGGTGTGACCTCA 541
Db 814 TGGTGATGAGACATCCCTTATGAGTGTGGAACGAGACCTTCAGCAGAGTGGTGTGACCTCA 873
QY 542 AGAAGCAGTGTGAGTGTGCTGCTGGAAGAGTTTGAAGAGTGTGAGCAGTGTGAGCAGTGTGACAGGA 601
Db 874 AGAANCAGTGTGATGTGCTGCTGGAAGAGTTTGAAGAGTGTGAGCAGTGTGAGCAGTGTGACAGGA 933
QY 602 ACCACGAGGAGGAGACCTGAGTGAATTC 630
Db 934 ACCACGAGGAGGAGACCTGAGTGAATTC 962

AC131887 246404 bp DNA linear HTG 09-MAY-2003
Rattus norvegicus clone CH230-3M14, WORKING DRAFT SEQUENCE, 3
unordered pieces.
AC131887
AC131887.3 GI:30466855
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 246404)
Muzny, D. Marie., Metaker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, W.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, S., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuwar, L., Louisied, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindrasekhar, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E.,
Mavhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plummer, P., Poindecker, A., Popovic, D., Primus, E., Pu, L. L.,
Puzo, M., Quilroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Roakey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shivatsbeyn, A., Sisson, I., Sitter, C. D., Smajda, D.,
Snead, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,

```



```
Db 24146 TTTGGTTGGTTCCTTTGGGTCCTTTCCAGCTCCAGAGTGGCCCTCCCATGTCCCTTTTC 24205
Qy 1592 TGCCTTTTGTCCCAATCCAGGG--GTTGGAAGGCCATCACCATCATTTGGAGGCTTAAC 1649
Db 24206 TGCCTATGTCCCAAGTCCCAAGTACTGGAAGGCTGTCACTGTCTCACTTGGAGGCTTAAC 24265
Qy 1650 GTGTCACTTACAGAGGTGC-----TGGAGCGCCCGGGGTTGGTTT 1692
Db 24266 CTGTAGTTACTAGGAGGACGAGTACTGGGAGCAGCGGGGGGGGGGGTGGTTG 24325
Qy 1693 GGGGTAATCACTCACTGTGCTCTCAGCTCTTAACACTGAGCCCTTAATACAGTTCTTT 1752
Db 24326 GAGTCAGTCACTCAGAGGTTCTCAGTCTTCTAACATGAGCCCTTAATACAGTTCTCT 24395
Qy 1753 CTGTTGTGTGACTCCCAAGCCCAACACACACATATAAATTAATTCGATGCTGTTTC 1812
Db 24386 ATGTTGTGTGACCCCATCCCT-----AAGCCATAAATTAATTTGTTGCTATTTC 24438
Qy 1813 ATAAGTGTAA 1822
Db 24439 CTAATGTAA 24448

RESULT 12
AR379595
LOCUS 1201 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 140 from patent US 6607879.
ACCESSION AR379595
VERSION AR379595.1 GI:40087229
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1201)
Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.
Compositions for the detection of blood cell and immunological
response gene expression
Patent: US 6607879-A 140 19-AUG-2003;
Incyte Corporation; Palo Alto, CA
JOURNAL
FEATURES
source
1..1201
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 25.4%; Score 467; DB 6; Length 1201;
Best Local Similarity 79.0%; Pred. No. 2.4e-103; Mismatches 95; Indels 68; Gaps 2;
Matches 612; Conservative 0;

Qy 228 GCTGAAGTCCGGCTTTTGGAGAAACGGGAAAGACCAAGGAAGTGAATGGACCGGCTATGG 287
Db 1 GCTGAAGTCCGGCTTTTGGAGAAACCGGCAAGACCAAGGAGTGAATGGACCGGCTATGG 60
Qy 288 CATCTGACGGGAAGGGCTCTGGAGTCAAGTACCAAGTCC----- 329
Db 61 CATCTGACCAAGAGGCTCTGGAGTCAAAATACAAAGTCCATTTTCAGATCCCCACAGA 120
Qy 330 -----GGACTTACGGTTAAT 345
Db 121 CCAGATGACCTATCTTCTTCAGCTCTGAGTCACTTCCCATTTGGGACTTGGGGTTAATC 180
Qy 346 GAAGTCACTGAGACCATTTTGAAGAGGCTTCTGGACTACAGCTTCGACCAAGGAGAGGACT 405
Db 181 GNAGTCACTGAGACCATTTTGAAGAGGCTTCTGGATTATAGCTTCGACCAAGGAGAGGACC 240
Qy 406 GCGACAAACCGGTTTCCCAAGGGTATGTGGAGACCTTTGAGACCTTCGACCAACCTAGTC 465
Db 241 GCGACCAATCGATTTTCCCAAGGGCATGTGAGAGACCTTTCAGAGACCTTTCAGACATTA 300
Qy 466 CACAAAGGGGTCAAGGTGTGATGATATCCCTATGAGCTGTGGAAAGAGAGCTTCAGCA 525
Db 301 CACAAAGGGGTCAAGGTGTGATGAGCATCCCTATGAGCTGTGGAAAGAGACTTCTGCA 360
```

```
Qy 526 GAGTGGCTGACCTCAAGAACGAGTGTGACGTGCTGGTGGAAAGATTGGAAGAGTGAATT 585
Db 361 GAGTGGCTGACCTCAAGAACGAGTGTGATGTGCTGGTGGAAAGATTGGAAGAGTGAATT 420
Qy 586 GAGGACTGTGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACCAAC 645
Db 421 GAGGACTGTGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACCAAC 480
Qy 646 GTGCTGAAGGAAAGGACACAGATTGCTTAGCAGAGCGGTGTCTTGGCAAGAGGGGGAC 705
Db 481 GTGCTGAAGGAAAGGACACAGATTGCTTAGCAGAGCAGTGTGTCGCGCAAGAGGGAGAC 540
Qy 706 ATAGCTCTCTGGGAGGGAAGAAATCCAAAGAACCCAGCGGAGTCAAG-----GGC 759
Db 541 ACAGCTGCTCTGGGAGGGAAGAAAGTCCAAAGAAAGAGCAGCAGCGGCCCAAGGACAGGC 600
Qy 760 TCCTCCAGTGCAGCAGCAAGCAGAGGAAGAACTGGGGGGGCTGGGGAGGATGCCAAC 819
Db 601 GCGAGGAGTAGCAGCAGCAAAACAAAGGAAGAGCTGGGTGGCTTGGAGGGAGACCCAGC 660
Qy 820 CCGGAGGAGGAGGGTGTGCAGAAAGCATCGCCCTCCACACAGACCCCTGATGAG 879
Db 661 CCGGAGGAGGATAGGGCATCCAGAAAGCATCCCTCTCACACACAGACCCCTGATGAG 720
Qy 880 CTGTGAGCCCAAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 721 CTCTGAGCCCAAGCATCTCTGTCTCTGAGACCCCTGATTTTGAAGCTGAGGA 775

RESULT 13
AR379729
LOCUS 1534 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 274 from patent US 6607879.
ACCESSION AR379729
VERSION AR379729.1 GI:40087363
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1534)
Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.
Compositions for the detection of blood cell and immunological
response gene expression
Patent: US 6607879-A 274 19-AUG-2003;
Incyte Corporation; Palo Alto, CA
JOURNAL
FEATURES
source
1..1534
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 24.9%; Score 458.6; DB 6; Length 1534;
Best Local Similarity 84.8%; Pred. No. 2.8e-101;
Matches 525; Conservative 0; Mismatches 92; Indels 2; Gaps 2;

Qy 5 CGAGGGAGGAAGCGCGCGGTCCGCTCTGCTCTGGTCCGGTGGGCCCATGGAGTCA 64
Db 660 CGCGGAGGAGGAACCGCGCGTCTTTAGGCTCGGGCCGCGCGGCATGGATTCAA 719
Qy 65 TGTCTGAGCTCGCGCCCGCTGCTCTATTTCCTTTGCTGCTGCTGCTTC-CGCTGCTG 123
Db 720 TGCCTGAGCCCGCTCGCGCTGTCTTCTGCTTCTTCTTCTGCTGCTGCTGCTGCTG 779
Qy 124 CTCCTCTCTCGCCGAGCTAGGCGCGGAGTCCCGCGGGGTGAGGAGACCGACTGGGTG 183
Db 780 NNNNNNNNNCCCGAGCTGGGCCGAGCAGCGCGGAGCTGAGGAGAACGACTGGGT 839
Qy 184 CGATTGCCAGCAAAATGCAAGTGTGCAAGTATGTTGCTGTGGAGCTG-AAAGTCGGCTTT 242
Db 840 CGCCTGCCAGCAAAATGCAAGTGTGTAATATGTTGCTGTGGAGCTGTAAAGTCAGCTT 899
Qy 243 TGAGGAACCGGGAAGAACCAAGGAAGTGAATGACACCGGCTATGGCATCTCGGACGGA 302
```

Db 900 TCAGAAACCGCAAGACCAAGAGGCTGATTGGCAGCGGCTATGTCATCTCGACCAAA 959
 Qy 303 GGGCTCTCGAGTCAAGTACCAAGTCGAGCTTACGGTTAATTGAAGTCAGTGAACCAT 362
 Db 960 GGCCTCTCGAGTCAAAATACCAAGTCGAGCTTGGGGTTAATCGAAGTCAGTGAACCAT 1019
 Qy 363 TTGCAAGAGGCTTCTGGACTACAGCCTCGACCAAGGAGAGGAGTGGCAGCAACCGGTTTGC 422
 Db 1020 TTGCAAGAGGCTTCTGGATTATAGCTGCAACAGGAGAGGAGCGCGCAGCAATCGATTGC 1079
 Qy 423 CAAGGGTATGCGGAGACTTTGAGACGCTGCAACCTAGTCCACAAAGGGGTCAAGGT 482
 Db 1080 CAAGGCCATGTCAGAGACTTTGAGACATTACAACTGTGTACAAAGGGGTCAAGGT 1139
 Qy 483 GGTGATGATATCCCTATGAGCTGTGGAACGAGACTTCTCGAGAGTGGCTGACCTCAA 542
 Db 1140 GGTGATGATATCCCTATGAGCTGTGGAACGAGACTTCTCGAGAGTGGCTGACCTCAA 1199
 Qy 543 GAAGCAGTGTGACGTGCTGGTGAAGAGTGTGAAGAGGTGATTGAGGACTGTTACAGGAA 602
 Db 1200 GAAGCAGTGTGATGTGCTGGTGAAGAGTGTGAGGAGTGTGAGGACTGTTACAGGAA 1259
 Qy 603 CCACCAGGAGGAGACCTG 621
 Db 1260 CCACCAGGAGGAGACCTG 1278

RESULT 14
 BC008133 LOCUS 1277 bp mRNA linear PRI 28-SEP-2004
 DEFINITION Homo sapiens cDNA clone IMAGE:3457829.
 ACCESSION BC008133
 VERSION BC008133.1 GI:14713265
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.
 1 (bases 1 to 1277)
 Strausberg,R.L., Feingold,B.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Hekim,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,H.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalek,U., Smalhus,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1277)
 DIRECTOR MGC Project.
 DIRECT SUBMISSION
 TITLE
 JOURNAL
 PUBLISHED
 REFERENCE
 AUTHORS
 Submitted (22-MAY-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgobcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: TRAK Plate: 19 Row: n Column: 9.

FEATURES

source

1..1277
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:3457829"
 /tissue_type="Cervix, carcinoma"
 /clone_lib="NIH_MGC_12"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 24.9%; Score 457.8; DB 8; Length 1277;
 Best Local Similarity 81.6%; Pred. No. 4.4e-101;
 Matches 544; Conservative 0; Mismatches 117; Indels 6; Gaps 1;
 Qy 274 GACACCGGCTATGCGATCCTGGAGGGAAGGGCTCTGGAGTCAAGTACACCAAGTCGGAC 333
 Db 173 GAGCCAGCGCGAGCTGAGGAGACGACTGGTTCGCTGCCAGCAATGCGAAGGGAC 232
 Qy 334 TTACGGTTAATGTAAGTCACTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCTGAC 393
 Db 233 TTGGCGTTAATCGAAGTCACTGAGACCATTTGCAAGAGGCTCCTGGATTATAGCTGAC 292
 Qy 394 AAGGAGGAGTGGCAGCAACCGGTTTGCACCAAGGATATGTCGAGACCTTTGAGACGCTG 453
 Db 293 AAGGAGGAGTGGCAGCAACCGGTTTGCACCAAGGATATGTCGAGACCTTTGAGACATTA 352
 Qy 454 CACAACCTAGTCCACAAAGGGGTCAAGTGGTGTGATGATATCCCTATGAGCTGTGGAAC 513
 Db 353 CACAACCTGGTACACAAAGGGGTCAAGTGGTGTGATGAGACATCCCTATGAGCTGTGGAAC 412
 Qy 514 GAGACCTCAGCAGAGTGGCTGACCTCAAGAGCAGTGTGACGTGCTGGTGGAGAGTTT 573
 Db 413 GAGACTTCTGACAGAGTGGCTGACCTCAAGAGCAGTGTGATGTGCTGGTGGAGAGTTT 472
 Qy 574 GAAGAGTGTATTGAGGACTGGTACAGGAACCAACAGGAGGAGACCTGACCTGAATTCCTC 633
 Db 473 GAGGAGTGTATTGAGGACTGGTACAGGAACCAACAGGAGGAGACCTGACCTGAATTCCTC 532
 Qy 634 TGTGCAACACCTGCTGAAGGGAAGGACACGAGTTCCCTAGCAGACGGTGGTCTGGC 693
 Db 533 TGCGCAACACCTGCTGAGGGAAGGACACGAGTTCCCTGGCAGACGAGTGGTCCGCG 592
 Qy 694 AAGAGGGGGACATAGCCTCCCTGGAGGGGAAGAAATCCAAAGAAAGCGCAGCGAGTTC 753
 Db 593 AAGAGGGGAGACACAGCTGCCCTGGAGGGGAAGAAATCCAAAGAAAGCGCAGCGAGTTC 652
 Qy 754 AAG-----GGCTCTCCAGTGGCAGCAGCAGCAGGAGGAGTGGGGGGCTCGGG 807
 Db 653 AAGCAGCAGCGCGGAGGAGTACAGCAAGCAAGGAGGAGTGGGGGGCTCGGG 712
 Qy 808 GAGGATGCCAACCGCGGAGGAGGAGTGTGCAAGAGGATCGCCCTCCACACAGC 867
 Db 713 GGAGACCCCGCGGAGGAGTACAGCAAGGAGTCCAGAGGATCCCTCTCACACACAGC 772
 Qy 868 CCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGTGATCAAGACCCCTGACTTCAGAG 927
 Db 773 CCCCTGTAGTGTGTGAGCCAGCTTGTGTCTTGTGATCAAGACCCCTGACTTCAGAG 832

Qy 928 CTTGGGA 934
 Db 833 CTGAGGA 839

RESULT 15

BC011767
 LOCUS Homo sapiens trinuclotide repeat containing 5, mRNA PFI 25-JUN-2004
 DEFINITION IMAGE:3609644, partial cds.
 ACCESSION BC011767
 VERSION BC011767.2 GI:40226374
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

REFERENCE 1 (bases 1 to 1593)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Alechul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932

TITLE

JOURNAL PubMed
 PUBLISHED 2 (bases 1 to 1593)
 DIRECT SUBMISSION

REFERENCE 2 (bases 1 to 1593)
 AUTHORS Strausberg, R.
 TITLE Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Dec 19, 2003 this sequence version replaced gi:15079962.
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 web site: <http://www.nhri.nih.gov/>

Contact: nisc.mgc@nhri.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brookes, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 27 Row: f Column: 20

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Location/Qualifiers
 source 1..1593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3609644"
 /tissue_type="Uterus, endometrium adenocarcinoma"
 /clone_lib="NIH_MGC_44"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

ORIGIN

Query Match 24.9%; Score 457.8; DB 8; Length 1593;
 Best Local Similarity 81.6%; Pred. No. 4.4e-101;
 Matches 544; Conservative 0; Mismatches 117; Indels 6; Gaps 1;
 Qy 274 GACACCGGCTATGGCATCTCGACGGGAGGGCTCTGGAGTCAAGTACACCAAGTCGGAC 333
 Db 475 GAGCCAGCCGGAGCTGAGGAGAACGACTGGGTTCCGCTGCCAGCAATGCGAAGGGAC 534
 Qy 334 TTACGGTTAATTGAAGTCACCTGAGACCAATTTCGAAGAGGCTTCTGGACTACAGCCTGCAC 393
 Db 535 TTGGCGTTAATCGAAGTCACCTGAGACCAATTTCGAAGAGGCTTCTGGATTATAGCCTGCAC 594
 Qy 394 AAGGAGGAGGACTGGCAGCAACCGGTTTCCCAAGGGTATGTCGGAGACCTTTGAGACGCTG 453
 Db 595 AAGGAGGAGGACCGGAGCAATCGATTTGCCAAGGGCATGTTCAGAGACCTTTGAGACATTA 654
 Qy 454 CACAACCTAGTCCCAAAAGGGGTCAAGTGTGTATGATATCCCTATGAGCTGTGGAAAC 513
 Db 655 CACAACCTGGTACAAAGGGGTCAAGTGTGTATGATGAGCATCCCCATATGAGCTGTGGAAAC 714
 Qy 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTGACGTGCTGGTGGAAAGAGTTT 573
 Db 715 GAGACTTCTGAGAGGTGGCTGACCTCAAGAGCAGTGTGATGTGCTGGTGGAAAGAGTTT 774
 Qy 574 GAAGAGGTGATTGAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTC 633
 Db 775 GAGAGGTGATCAGGAGCTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTC 834
 Qy 634 TGTCCCAACCAAGCTGTGTAAGGGAAGAGACACGAGTTGCTTACAGAGAGCGGTGTGCTGGC 693
 Db 835 TGCGCCAACCAAGCTGTGTAAGGGAAGAGACACCAAGTTGCTTGGCAGAGCAGTGTGCTGGC 894
 Qy 694 AAGAGGGGACATAGCTCCCTGGAGGGAAGAAATCCAAAGAGAGAGAGAGAGAGAGAGAGAG 954
 Db 895 AAGAGGGGAGACACAGCTGCTGGAGGGAAGAGAGTCCAAAGAGAGAGAGAGAGAGAGAGAG 954
 Qy 954 AAG-----GGCTCTCCAGTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 807
 Db 955 AAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1014
 Qy 808 GAGAGTCACCAAGCGCAG 867
 Db 1015 GGAGACCCCAAGCCCGCAGAGAGAGATCCAGAGAGAGAGATCCCAAGAGAGAGATCCCAAGAG 1074
 Qy 868 CCCCTGATGAGCTGTGAGCCCGCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCACAG 927
 Db 1075 CCCCTGATGAGCTGTGAGCCCGCAGCTTAGTGTCTTGAATCAAGACCCCTGACTTCACAG 1134
 Qy 928 CTTGGGA 934
 Db 1135 CTGAGGA 1141

Search completed: December 30, 2005, 01:45:58
 Job time : 9087 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 19:26:35 ; Search time 1051 Seconds
(without alignments)
11661.627 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*
- 14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1839	100.0	1839	4	AAS05278
2	1725.2	93.8	1848	4	AAS05286
3	1528	83.1	1553	14	ADV67441
4	691.6	37.6	1378	10	ACA54677
5	691.6	37.6	1378	14	ADUB3175
6	691.6	37.6	1675	8	AAS21311
7	691.6	37.6	1675	8	ACA03670
8	691.6	37.6	1675	8	ABX89208
9	691.6	37.6	1675	8	ACD41862
10	691.6	37.6	1675	8	ACA04091
11	691.6	37.6	1675	9	ADA45654
12	691.6	37.6	1675	9	ADA76085
13	691.6	37.6	1675	9	ADA18735
14	691.6	37.6	1675	9	ADA61358
15	691.6	37.6	1675	9	ADB19143
16	691.6	37.6	1675	9	ADB27684
17	691.6	37.6	1675	9	ADA86163
18	691.6	37.6	1675	9	ADB15727
19	691.6	37.6	1675	9	ADA47513

20	691.6	37.6	1675	9	ADA67308	Ada67308 Human PRO
21	691.6	37.6	1675	9	ADB30315	Adb30315 CDNA enco
22	691.6	37.6	1675	9	ADA85611	Ada85611 Novel hum
23	691.6	37.6	1675	9	ADA96823	Ada96823 Human PRO
24	691.6	37.6	1675	9	ADA79127	Ada79127 Human PRO
25	691.6	37.6	1675	9	ADA87266	Ada87266 Novel hum
26	691.6	37.6	1675	9	ADB16468	Adb16468 Human PRO
27	691.6	37.6	1675	9	ADA91560	Ada91560 Novel hum
28	691.6	37.6	1675	9	ADB14623	Adb14623 Human PRO
29	691.6	37.6	1675	9	ADB18584	Adb18584 Novel hum
30	691.6	37.6	1675	9	ADA93799	Ada93799 Human PRO
31	691.6	37.6	1675	9	ADB19695	Adb19695 Novel hum
32	691.6	37.6	1675	9	ADB13007	Adb13007 Human PRO
33	691.6	37.6	1675	9	ACD98491	AcD98491 Novel hum
34	691.6	37.6	1675	9	ADA74261	Ada74261 Human PRO
35	691.6	37.6	1675	9	ADB24494	Adb24494 Human PRO
36	691.6	37.6	1675	9	ADA82018	Ada82018 Human PRO
37	691.6	37.6	1675	9	ADA74981	Ada74981 Human PRO
38	691.6	37.6	1675	9	ADA85059	Ada85059 Novel hum
39	691.6	37.6	1675	9	ADA84507	Ada84507 Novel hum
40	691.6	37.6	1675	9	ADB29763	Adb29763 CDNA enco
41	691.6	37.6	1675	9	ADA80291	Ada80291 Human PRO
42	691.6	37.6	1675	9	ADA75533	Ada75533 Human PRO
43	691.6	37.6	1675	9	ADA46758	Ada46758 Human PRO
44	691.6	37.6	1675	9	ADB25054	Adb25054 Human PRO
45	691.6	37.6	1675	9	ADA93230	Ada93230 Human PRO

ALIGNMENTS

RESULT 1
AAS05278

ID AAS05278 standard; cDNA; 1839 BP.

XX AAS05278;

AC AAS05278;

XX 07-SEP-2001 (first entry)

XX Murine trinucleotide repeat protein (TRP) cDNA sequence.
XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX transgenic animal; knockout mouse; triplet repeat expansion;
XX fragile X syndrome; Huntington's disease; mouse; murine; ss.

OS Mus sp.

XX Key Location/Qualifiers

XX CDS 55..885

XX /*tag= a

XX /product= "TRP"

XX WO200130798-A1.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-US029382.

XX PR 26-OCT-1999; 99US-0161488P.

XX PA (DELTA-) DELTAGEN INC.

XX PI Klein R, Matthews W, Moore M, Allen KD;

XX DR WPI; 2001-300473/31.

XX DR P-ESDB; AAU02498.

XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP.

XX Claim 8; Fig 11; 106pp; English.

CC The present sequence encodes for murine trinucleotide repeat protein
 CC (TRP). The invention describes methods of producing embryonic stem (ES)
 CC cells comprising a heterozygous disruption in a target DNA sequence
 CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
 CC comprising a homozygous disruption in a gene encoding TRP, where the
 CC disruption inhibits the production of the wild type TRP. The invention
 CC also relates to identifying agents capable of affecting a phenotype of a
 CC knockout mouse. Also described are methods of determining whether
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
 CC phenotypic change. The transgenic animals and the cells are useful for
 CC identifying compounds capable of ameliorating disease symptoms, and as
 CC test substrates for the identification of drugs, pharmaceuticals,
 CC therapies and interventions which may be effective in treating
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
 CC disease. The animal models for trinucleotide repeat disorders are ideal
 CC model systems to study the progression of disease in vivo, the molecular
 CC basis of these diseases and show the features observed in human disease.
 CC Using the mice, it is possible to model both the pathogenic mechanism and
 CC the trinucleotide repeat instability in the mouse

XX
 Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 U; 0 Other;

Query Match 100.0%; Score 1839; DB 4; Length 1839;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGAGGAGGAGCGCCGGGTCGGCTCTGCTCTGGGTCCGGTGGCCATGGAG 60
 DB 1 GGCACGAGGAGGAGCGCCGGGTCGGCTCTGCTCTGGGTCCGGTGGCCATGGAG 60
 QY 61 TCCATGCTGAGCTCGCCCGCTGCTCTTATTTCTTTGCTGCTGCTCTCCGCTG 120
 DB 61 TCCATGCTGAGCTCGCCCGCTGCTCTTATTTCTTTGCTGCTGCTCTCCGCTG 120
 QY 121 CTGCTCTCTCCCGAAGCTAGGCCCGAGTCCCGCGGGCTGAGGACCGAATGG 180
 DB 121 CTGCTCTCTCCCGAAGCTAGGCCCGAGTCCCGCGGGCTGAGGACCGAATGG 180
 QY 181 GTGGATTTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGGAGCTGAACTCGGCT 240
 DB 181 GTGGATTTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGGAGCTGAACTCGGCT 240
 QY 241 TTTGAGGAAAACGGGAAACCAAGAAAGTGTGACACCGGTATGGCATCTTGACGCGG 300
 DB 241 TTTGAGGAAAACGGGAAACCAAGAAAGTGTGACACCGGTATGGCATCTTGACGCGG 300
 QY 301 AAGGCTCTGGAGTCAAGTACACCAAGTCCGACTTACGGTTAATTGAAGTCACTGAGACC 360
 DB 301 AAGGCTCTGGAGTCAAGTACACCAAGTCCGACTTACGGTTAATTGAAGTCACTGAGACC 360
 QY 361 ATTTGCAAGAGGCTTCTGGACTACAGCTGACAGGAGGAGCTGGCAGCAACCGGTTT 420
 DB 361 ATTTGCAAGAGGCTTCTGGACTACAGCTGACAGGAGGAGCTGGCAGCAACCGGTTT 420
 QY 421 GCCAAGGATGTCTGGAGACCTTTGAGACGCTGACCAACCTAGTCCACAAAGGGGTCAAG 480
 DB 421 GCCAAGGATGTCTGGAGACCTTTGAGACGCTGACCAACCTAGTCCACAAAGGGGTCAAG 480
 QY 481 GTGGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGAGTGGCTGACCTC 540
 DB 481 GTGGTGTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGAGTGGCTGACCTC 540
 QY 541 AAGAAGCATGTGACGCTGCTGGTGAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG 600
 DB 541 AAGAAGCATGTGACGCTGCTGGTGAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG 600
 QY 601 AACCAACGAGGAGAGACTGACTGAAATTTCTCTGTGCAACCAACCTAGTCCACAAAGGGGTCAAG 660
 DB 601 AACCAACGAGGAGAGACTGACTGAAATTTCTCTGTGCAACCAACCTAGTCCACAAAGGGGTCAAG 660
 QY 661 GACACGAGTGTGCTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGGAATAGCCCTCCCTGGGA 720
 DB 661 GACACGAGTGTGCTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGGAATAGCCCTCCCTGGGA 720

QY 721 GGGAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAAG 780
 DB 721 GGGAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCCTCCAGTGGCAGCAGCAAG 780
 QY 781 CAGAGGAAGAACTGGGGGGCTTGGGGAGATGCCAAACGCCGAGGAGGAGAGGGTGTG 840
 DB 781 CAGAGGAAGAACTGGGGGGCTTGGGGAGATGCCAAACGCCGAGGAGGAGAGGGTGTG 840
 QY 841 CAGAGGCATCGCCCTCCACACAGCCCTCATGAGCTGTGAGCTGTGAGCCACTTAGTGTG 900
 DB 841 CAGAGGCATCGCCCTCCACACAGCCCTCATGAGCTGTGAGCTGTGAGCCACTTAGTGTG 900
 QY 901 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCT 960
 DB 901 CTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCT 960
 QY 961 CCAGCAAGGACAGCTGTGTCAGCATCAGGTCTCTCCCTTGGGTGTGCCCCCTTTCTT 1020
 DB 961 CCAGCAAGGACAGCTGTGTCAGCATCAGGTCTCTCCCTTGGGTGTGCCCCCTTTCTT 1020
 QY 1021 CCCTTGAACACAGAGGTGGAAGATCTGGGTGCTGGGAGCGCACCCCAAAGG 1080
 DB 1021 CCCTTGAACACAGAGGTGGAAGATCTGGGTGCTGGGAGCGCACCCCAAAGG 1080
 QY 1081 GAAGAGGAGGAGGAGCAGAGCGAGCTCTCTTCTTCAACAGTCCCTTCAAGAGCTCCGG 1140
 DB 1081 GAAGAGGAGGAGGAGCAGAGCGAGCTCTCTTCTTCAACAGTCCCTTCAAGAGCTCCGG 1140
 QY 1141 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATG 1200
 DB 1141 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATG 1200
 QY 1201 AGGCACATAGATGCTCACACGCGCCCTTCAAGCAGGAGGAGTCCGTGCGAGCTCAG 1260
 DB 1201 AGGCACATAGATGCTCACACGCGCCCTTCAAGCAGGAGGAGTCCGTGCGAGCTCAG 1260
 QY 1261 CAGCAGGCTGCTCTCTTCCACCAAGCATTTCTTCTGCTGGTCTTGTGCGATGG 1320
 DB 1261 CAGCAGGCTGCTCTCTTCCACCAAGCATTTCTTCTGCTGGTCTTGTGCGATGG 1320
 QY 1321 TAAATTCCAGAACTTCCAGGACAAACTCGGTGTGGCAAAAAGGGGTGGAAGCAGAGC 1380
 DB 1321 TAAATTCCAGAACTTCCAGGACAAACTCGGTGTGGCAAAAAGGGGTGGAAGCAGAGC 1380
 QY 1381 CAGAGCCACGACAGACTGACAGAGGCGACCTGACCTTAACCCCTTGGAAAGCCAAATC 1440
 DB 1381 CAGAGCCACGACAGACTGACAGAGGCGACCTGACCTTAACCCCTTGGAAAGCCAAATC 1440
 QY 1441 TGCAAGTTCCTGTCACCCACTCTCTGAGGAGCGCTCATGCTCTGCCCAGGCTTCT 1500
 DB 1441 TGCAAGTTCCTGTCACCCACTCTCTGAGGAGCGCTCATGCTCTGCCCAGGCTTCT 1500
 QY 1501 CCCAGGGTACACAGATAAACACACTTTGGCCTTTGGCTTTGGTCTCTGGGTCTCATCA 1560
 DB 1501 CCCAGGGTACACAGATAAACACACTTTGGCCTTTGGCTTTGGTCTCTGGGTCTCATCA 1560
 QY 1561 GCCTCCAGAGTGTCCCTCATGATCTTTTGGCTTTGTCCTCCCAATCCAGGGGCTGG 1620
 DB 1561 GCCTCCAGAGTGTCCCTCATGATCTTTTGGCTTTGTCCTCCCAATCCAGGGGCTGG 1620
 QY 1621 AAGGCCATCACCATCATTTGGAGGCTTAACCTGTAGTGTACAGTGTGAGGAGCGCC 1680
 DB 1621 AAGGCCATCACCATCATTTGGAGGCTTAACCTGTAGTGTACAGTGTGAGGAGCGCC 1680
 QY 1681 CGGGGTTGGTTGGGGTAATCACTCAGTGGCTCTCAGCTTCTAACACTGAGGCCCTTA 1740
 DB 1681 CGGGGTTGGTTGGGGTAATCACTCAGTGGCTCTCAGCTTCTAACACTGAGGCCCTTA 1740
 QY 1741 ATACAGTTCCTTGTGTGTGTGAGTCCACGCGCCCAACACACACACATATAATTT 1800
 DB 1741 ATACAGTTCCTTGTGTGTGTGAGTCCACGCGCCCAACACACACATATAATTT 1800

Query Match 93.8%; Score 1725.2; DB 4; Length 1848;
Best Local Similarity 96.6%; Pred. No. 0;

566	DB		GTGTGACGTGCTGGTGGAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCA	625
609	QY		GGAGGAAGACCTGACTCAATTCCTCTGTGTGCCAACACAGCTGCTGAAGAGGAAGACACGAG	668
626	DB		GGAGGAAGACCTGACTGAATTCCTCTGTGTGCCAACACAGCTGCTGAAGAGGAAGACACGAG	685
669	QY		TTGCTCTAGCAGAGCGGTGTCTGGCAAGAAGGGGGACATAGCTCCCTGGAGGGGAAGAA	728
686	DB		TTGCTCTAGCAGAGCGGTGTCTGGCAAGAAGGGGGACATAGCTCCCTGGAGGGGAAGAA	745
729	QY		ATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAA	788
746	DB		ATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAA	805
789	QY		GGAATCTGGGGGCTTGGGGAGGATGCCAACCGCGAGGAGGAGGGTGTGCAGAAGGC	848
806	DB		GGAATCTGGGGGCTTGGGGAGGATGCCAACCGCGAGGAGGAGGGTGTGCAGAAGGC	865
849	QY		ATCGCCCTCCCAACAGACCCCTCTGATGAGCTGTGAGCCACGCTTAGTCTCTTGAATC	908
866	DB		ATCGCCCTCCCAACAGACCCCTCTGATGAGCTGTGAGCCACGCTTAGTCTCTTGAATC	925
909	QY		AAGACCCCTGACTTCAGAGCTTGGGACACGCACAGCGCAGCGCAGCTCCAGCAAG	968
926	DB		AAGACCCCTGACTTCAGAGCTTGGGACACGCACAGCGCAGCGCAGCTCCAGCAAG	985
969	QY		GACAGCTGTGTCCAGCATCAGTCTCTCCCTTGGCTGTGCCCTTTCCTTCCCTTGAA	1028
986	DB		GACAGCTGTGTCCAGCATCAGTCTCTCCCTTGGCTGTGCCCTTTCCTTCCCTTGAA	1045
1029	QY		CAACAGCAAGAGTGGGAAGGATCTGGGGTCTCTGGGAGACGGCACCCCAAGGGAAGGAA	1088
1046	DB		CAACAGCAAGAGTGGGAAGGATCTGGGGTCTCTGGGAGACGGCACCCCAAGGGAAGGAA	1105
1089	QY		GGAGGAGCAAGAGGAGCTCTTTTCTACAGCTGCCCTCTACAGGCTCGGGGTTCACC	1148
1106	DB		GGAGGAGCAAGAGGAGCTCTTTTCTACAGCTGCCCTCTACAGGCTCGGGGTTCACC	1165
1149	QY		CAGCATCCCCAGGCTGAGATCTCAGGCTCTCTGACATGGAAGCTCAAGAGCATGAGGCACAT	1208
1166	DB		CAGCATCCCCAGGCTGAGATCTCAGGCTCTCTGACATGGAAGCTCAAGAGCATGAGGCACAT	1225
1209	QY		AAGATGCTCACAGCGCCCTTTCAGCCAGGAAGGACTCCGTGGAGCCTCAGCAGCCAGG	1268
1226	DB		AAGATGCTCACAGCGCCCTTTCAGCCAGGAAGGACTCCGTGGAGCCTCAGCAGCCAGG	1285
1269	QY		CGTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGGATGGTAAATTCG	1328
1286	DB		CGTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGGTCTTGTGGATGGTAAATTCG	1345
1329	QY		AGAACTTCCAGGACAACTCGGGTGTGGCACAAAGGGGCTGGAGCCGCAGAGCCAGAGCCA	1388
1346	DB		AGAACTTCCAGGACAACTCGGGTGTGGCACAAAGGGGCTGGAGCCGCAGAGCCAGAGCCA	1405
1389	QY		CGCCAGAGACTGCAGAGAGGGCAGCTGACCTTAACCCCTCTGGAAAGGCCAATCTGCAGTTC	1448
1406	DB		CGCCAGAGACTGCAGAGAGGGCAGCTGACCTTAACCCCTCTGGAAAGGCCAATCTGCAGTTC	1465
1449	QY		CCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGTCTGCCACAGCCCTTCTCCAGGGC	1508
1466	DB		CCGTGTCCACCCACTCTCTCTGAGGAGCGCTCATGTCTGCCACAGCCCTTCTCCAGGGC	1525
1509	QY		TACCAGAGTAAACACCTTTTGGCCTTTC	1536
1526	DB		TACCAGAGTAAACACCTTTTGGCCTTTC	1553

RESULT 4
ACA54677
ID ACA5
XX
AC ACA5

[illegible]

05-JUN-2003 (first entry)
Human NF-kappaB associated polynucleotide sequence #117.

Human, nuclear factor-kappaB, NF-kappaB; immune disorder; cancer; inflammatory disorder; apoptosis; hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour; hyper-IgM syndrome; viral infection; asthma; hypohidrotic ectodermal dysplasia; human immunodeficiency virus; HIV; X-linked anhidrotic ectodermal dysplasia; al incontinentia pigmenti; influenza; rheumatoid arthritis; inflammatory bowel disease; colitis; atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE; experimental allergic encephalomyelitis; autoimmune disorder; wound; hyper immune activity; acute phase response; hypercongenital condition; birth defect; necrotic lesion; organ transplant rejection; pancreas; signal transduction; hyperproliferative disorder; diabetes mellitus; vitamin B12 malabsorption; neurological disorder; Huntington's chorea; Turner's syndrome; bacterial infection; cardiovascular disorder; infertility; porriasis; haemolytic anaemia; antiinflammatory; anti-HIV; cycostatic; hepatotropic; virucide; antirheumatic; antiarthritic; antiasthmatic; immunomodulator; antidiabetic; antiallergic; neuroprotective; immunosuppressive; vulnerary; antibacterial; antiinfertility; antihaemic; antipsoriatic; cerebroprotective; cardiac; antiarteriosclerotic; gene; ss.

Homo sapiens.

WO200286076-A2.

31-OCT-2002.

19-APR-2002; 2002WO-US012636.

19-APR-2001; 2001US-0284962P.

28-AFK-2001; 2001US-028804JE;
09-JAN-2002; 2002US-0346986P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Carman J, Feder J, Nadler S;

WPI; 2003-093119/08.

P-PSDB; ABU69616.

Novel NF- κ B-associated polypeptides and polynucleotides useful for diagnosing, treating and preventing cancer, hepatic disorders, aberrant apoptosis, viral infections, autoimmune disorders, asthma and stroke.

Claim 4: Fig 10H; 608pp; English.

The present invention relates to the isolation of human nuclear factor-kappaB (NF-kappaB) associated polypeptides and polynucleotides. The NF-kappaB associated polypeptide and polynucleotide sequences are useful for preventing, treating or ameliorating various disorders including immune disorders, inflammatory disorders, cancers, disorders relating to aberrant apoptosis, hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic ectodermal dysplasia, immunodeficiency, albinism, incontinentia pigmenti, viral infections (e.g. those caused by human immunodeficiency virus (HIV), human T-cell lymphotropic virus (HTLV), hepatitis B, hepatitis C, Epstein Barr virus (EBV), influenza), rheumatoid arthritis, inflammatory bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick syndrome, stroke, experimental allergic encephalomyelitis (EAE), autoimmune disorders, disorders related to hyper immune activity, disorders related to aberrant acute phase responses, hypercongenital conditions, birth defects, necrotic lesions, wounds, organ transplant rejection, disorders related to aberrant signal transduction, hyperproliferative disorders, diseases of the pancreas (e.g. diabetes mellitus, vitamin B12 malabsorption), neurological disorders (e.g. Huntington's chorea), Turner's syndrome, bacterial infections, cardiovascular disorders, infertility, psoriasis and haemolytic anaemia. The present sequence represents a human NF-kappaB associated polynucleotide of the invention.

XX	SQ	Sequence	1378 BP; 311 A; 399 C; 419 G; 249 T; 0 U; 0 Other;
		Query Match	37.6%; Score 691.6; DB 10; Length 1378;
		Best Local Similarity	84.5%; Pred. No. 6.4e-168;
		Matches	791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
Qy	5	CGAGGGAGGAAGCGCCGCGGGTCCGCTCTGCTCTGGGTCCGGCTGCGGCACATGAGTCCA	64
Db	21	CGCGGAGGAGGAACCGCCCGTCTTTAGGGTCCGGGCCCGCGGCATGATTCAA	80
Qy	65	TGCTGAGCTCGCGCCCGCTCTTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	124
Db	81	TGCTGAGCCGCGTCCGCTGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG	140
Qy	125	TCCTTCTCCGCCCAAGCTAGGCCCGAGTCCCGCCGGGCTGAGGAGACCGACTGGGTGC	184
Db	141	TGCTGCCGCCCCCGAGCTGGCCCGAGCCAGCCGGGCTGAGGAGAACGACTGGGTTC	200
Qy	185	GATTGCCAGCAATATGCAAGTGTGCAATGATATGTTGCTGTGGAGCTGAAFTCGGCTTTG	244
Db	201	GCCTGCCAGCAATATGCAAGTGTGTAATATGTTGCTGTGGAGCTGAAFTCAGCCTTTG	260
Qy	245	AGGAAACGGGAAGACCAAGAAAGTATGACACGGCTATGGCATCTCTGGACGGGAGG	304
Db	261	AGGAAACGGGAAGACCAAGAGAGTATGAGCATCTCTGGACGGGCTATGGCATCTCTGGACGGGAGG	320
Qy	305	GCTCTGGAGTCAAGTATACCAAGTCCGAGTTCACGGTTAAATGAAAGTCACTGAGACCAATTT	364
Db	321	CCTCTGGAGTCAATATACCAAGTCCGAGTTCACGGTTAAATGAAAGTCACTGAGACCAATTT	380
Qy	365	GCAAGAGCTTCTGGACTACAGCTTCAACCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCA	424
Db	381	GCAAGAGCTTCTGGAATATAGCTTCAACCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCA	440
Qy	425	AGGCTATGTCGGAGACCTTTGAGAGCTGCAACACTAGTCCACAAAGGGGTCAAGTGG	484
Db	441	AGGCAATGTCAGAGACCTTTGAGACATTCACAACTGGTATACAAAGGGGTCAAGTGG	500
Qy	485	TGATGGATATCCCTATGAGCTGTGGAAACGAGACTCAGCAGAGGTGGCTGACCTCAAGA	544
Db	501	TGATGGATATCCCTATGAGCTGTGGAAACGAGACTCCTGACAGAGGTGGCTGACCTCAAGA	560
Qy	545	AGCAGTGTGCTGCTGGTGAAGAGTTTGAAGAGGTGATTTGAGGACTGTTGACAGGAACC	604
Db	561	AGCAGTGTGATGTGCTGTGAAGAGTTTGAAGAGGTGATTCAGAGACTGTTGACAGGAACC	620
Qy	605	ACCAAGGAGGAGACTGACCTAAATTCCTCTGTGCCAACCACTGCTGAAAGGAAAGGACA	664
Db	621	ACCAAGGAGGAGACTGACCTAAATTCCTCTGTGCCAACCACTGCTGAAAGGAAAGGACA	680
Qy	665	CGAGTGTGCTAGCAGACGGTGGTCTGGCAAGAGGGGGAATAGCTTCCTCGGAGGGGA	724
Db	681	CCAGTGTGCTGGCAGACAGTGGTTCGGCAGAGAGGGAGACACAGCTGCTCTGGAGGGGA	740
Qy	725	AGAAATCCAAAGAAAGCGCAGCGAGTCAAG-----GGCTCTCTCAGTGGCAGCAGCA	778
Db	741	AGAAATCCAAAGAAAGCAGCAGGCGCAAGGCAAGCGGCGGAGGATAGCAGCAGCA	800
Qy	779	AGCAGAGGAGAACTGGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGGGTG	838
Db	801	AACAAAGGAGAGGAGCTGGTGGCTCTGAGGGAGAGCCCCAGCCCCGAGAGAGATGAGGGA	860
Qy	839	TGCAGAAAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCCAGCTTAGTG	898
Db	861	TCCAGAGGCATCCCTCTCACACAGCCCCCTGATGAGCTGTGAGCCCCACCCAGCAT	920
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAAGAGCTTGGGA	934
Db	921	CCTCTGTCTCTGAGACCCCTGATTTTGAAGCTGAGGA	956

RESULT 5

2

CC propagation in cells infected with other viruses, associated with EL-8,
 CC disorders associated with aberrant IL-8 expression, disorders associated
 CC with aberrant IL-8 activity, pulmonary disorders, pulmonary fibrosis,
 CC Behcet's disease, bacterial infections, gynecological diseases,
 CC psoriasis, IgA nephropathy, chronic obstructive pulmonary disease,
 CC Kawasaki disease, Crohn's disease, peripheral arterial occlusive disease,
 CC Hodgkin's disease, idiopathic intermediate uveitis, hyaline membrane
 CC disease, acute rheumatic fever, chronic rheumatic heart disease,
 CC ulcerative colitis, autoimmune disorders, and autoimmune thyroid disease.
 CC The present sequence encodes a protein found to interact with the NFkappa
 CC B modulating protein AD037.

XX Sequence 1378 BP; 311 A; 399 C; 419 G; 249 T; 0 U; 0 Other;

Query Match 37.6%; Score 691.6; DB 14; Length 1378;
 Best Local Similarity 84.5%; Pred. No. 6.4e-168;
 Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY	5	CGAGGAGGAAGCGCGCGGTCGCTCTGCTCTGGGTCCGGCTGGGCGCATGGAGTCCA	64
DB	21	CGCGGAGGAGGAACCGCGCGTCTTAGGGTCCGGCGCGCGGCCCATGGATTCAA	80
QY	65	TGCTGAGCTCGCGCGCGCTGCTCTTATTTCTTGTCTGCTGCTCGCTGCTGCTG	124
DB	81	TGCCTGAGCGCGCGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTGCTG	140
QY	125	TCCTTCCTCCCGGAGCTAGCGCGGCTGCGCGCGGCTGAGGAGACCGACTGGGTGC	184
DB	141	TCCTCGCGCGCGGAGCTGGCGCGCGGAGCGCGGAGCTGAGGAGAACGACTGGGTTC	200
QY	185	GATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTG	244
DB	201	GCTTCCCGCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCTTTG	260
QY	245	AGGAACCGGAAGACCAAGGAGTGTGACACCGCTATGGCATCTGACCGGAGG	304
DB	261	AGGAACCGGCAAGACCAAGGAGTGTGACACCGCTATGGCATCTGACCGGAGG	320
QY	305	GCTCTGAGTCAAGTACACCAAGTGTGACACCGCTATGGCATCTGACCGGAGG	364
DB	321	CCTCTGAGTCAAAATACCAAGTGTGACACCGCTATGGCATCTGACCGGAGG	380
QY	365	GCAAGAGCTTCTGAGTACAGCTGTGCAAGGAGGAGTGTGCAAGGAGGAGTGTG	424
DB	381	GCAAGAGCTTCTGAGTATAGCTGTGCAAGGAGGAGTGTGCAAGGAGGAGTGTG	440
QY	425	AGGATGTCGGAGACCTTGTGAGCTGTGCAAGGAGGAGTGTGCAAGGAGGAGTGTG	484
DB	441	AGGATGTCGGAGACCTTGTGAGCTGTGCAAGGAGGAGTGTGCAAGGAGGAGTGTG	500
QY	485	TGATGGATATCCCTATGAGTGTGGAAGGAGGAGTGTGCAAGGAGGAGTGTGCAAG	544
DB	501	TGATGGATATCCCTATGAGTGTGGAAGGAGGAGTGTGCAAGGAGGAGTGTGCAAG	560
QY	545	AGCAGTGTGAGTGTGTTGGAAGGAGTGTGGAAGGAGGAGTGTGGAAGGAGGAGTGTG	604
DB	561	AGCAGTGTGAGTGTGTTGGAAGGAGTGTGGAAGGAGGAGTGTGGAAGGAGGAGTGTG	620
QY	605	ACCAGGAGGAGACCTGTGAGTGTGTTGGAAGGAGTGTGGAAGGAGGAGTGTGGAAG	664
DB	621	ACCAGGAGGAGACCTGTGAGTGTGTTGGAAGGAGTGTGGAAGGAGGAGTGTGGAAG	680
QY	665	CGAGTTGCTGTGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAG	724
DB	681	CGAGTTGCTGTGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAG	740
QY	725	AGAAATCCAAAGAGGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAG	778
DB	741	AGAAATCCAAAGAGGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAG	800
QY	779	AGCAGAGGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAGGAGGAGTGTG	838
DB	801	AACAAGGAGGAGGAGTGTGTTGGAAGGAGGAGTGTGGAAGGAGGAGTGTGGAAGGAGGAGTGTG	860

QY	839	TGCAGAGGATCCGCCCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTG	898
DB	861	TCCAGAGGATCCGCCCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCAGCTTAGTG	920
QY	899	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA	934
DB	921	CCTCTGCTCTGAGACCCCTGATTTTGAAGCTGAGGA	956

RESULT 6

AAS21311
 ID AAS21311 standard; cDNA; 1675 BP.

AC AAS21311;
 XX

DT 24-OCT-2001 (first entry)
 XX

DE Human cDNA sequence encoding for PRO4409 polypeptide.
 XX

Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast;
 prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage;
 ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte;
 A-peptide; factor VIIA; gene therapy; ss.

OS Homo sapiens.
 XX

PN WO200140466-A2.
 XX

PD 07-JUN-2001.
 XX

PF 01-DEC-2000; 2000WO-US032678.
 XX

PR 01-DEC-1999; 99WO-US028301.
 PR 01-DEC-1999; 99WO-US028634.
 PR 02-DEC-1999; 99WO-US028551.
 PR 02-DEC-1999; 99WO-US028564.
 PR 02-DEC-1999; 99WO-US028565.
 PR 09-DEC-1999; 99US-0170262P.
 PR 16-DEC-1999; 99WO-US030095.
 PR 20-DEC-1999; 99WO-US030911.
 PR 20-DEC-1999; 99WO-US030999.
 PR 30-DEC-1999; 99WO-US031243.
 PR 30-DEC-1999; 99WO-US031274.
 PR 06-JAN-2000; 2000WO-US000219.
 PR 06-JAN-2000; 2000WO-US000277.
 PR 06-JAN-2000; 2000WO-US003376.
 PR 11-FEB-2000; 2000WO-US003565.
 PR 18-FEB-2000; 2000WO-US004341.
 PR 22-FEB-2000; 2000WO-US004414.
 PR 24-FEB-2000; 2000WO-US004914.
 PR 24-FEB-2000; 2000WO-US005004.
 PR 01-MAR-2000; 2000WO-US005601.
 PR 02-MAR-2000; 2000WO-US005841.
 PR 03-MAR-2000; 2000US-0187202P.
 PR 10-MAR-2000; 2000WO-US006319.
 PR 15-MAR-2000; 2000WO-US006884.
 PR 20-MAR-2000; 2000WO-US007377.
 PR 21-MAR-2000; 2000WO-US007532.
 PR 30-MAR-2000; 2000WO-US008439.
 PR 17-MAY-2000; 2000WO-US013705.
 PR 22-MAY-2000; 2000WO-US014042.
 PR 30-MAY-2000; 2000WO-US014941.
 PR 02-JUN-2000; 2000WO-US015264.
 PR 05-JUN-2000; 2000US-0209832P.
 PR 28-JUL-2000; 2000WO-US020710.
 PR 11-AUG-2000; 2000WO-US022031.
 PR 23-AUG-2000; 2000WO-US023522.
 PR 24-AUG-2000; 2000WO-US023328.
 PR 08-NOV-2000; 2000WO-US030952.
 PR 10-NOV-2000; 2000WO-US030873.
 XX

PA	(GETH) GENENTECH INC.	Db	707	AGGCATGTCTAGAGACCTTTGAGACATTTACACAACTGGGTACACAAAGGGGTCAAGGTGG	766
XX					
PI	Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;	Qy	485	TGATGGATATCCCTTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA	544
PI	Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;	Db	767	TGATGGACATCCCTTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826
XX	Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;				
DR	WPI; 2001-408281/43.	Qy	545	AGCAGTGTGACGTGCTGGTGGAGAGTTTGAAGAGGTGATTGAGGACTGTGGTACAGGAACC	604
DR	P-PSDB; AAU12239.	Db	827	AGCAGTGTGATGTGCTGGTGGAGAGTTTGAAGAGGTGATCGAGGACTGTGTACAGGAACC	886
XX					
PT	Isolated, secretory and transmembrane PRO polypeptide used to detect	Qy	605	ACGAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAGGAGGAAGGACA	664
PT	other PRO polypeptides, link bioactive molecules to cells expressing PRO	Db	887	ACGAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACCACTGCTGAGGAGGAAGGACA	946
PT	polypeptides, and detect the presence of mammalian tumors e.g. lung,				
XX	breast, prostate, cervical.				
XX					
PS	Claim 3; Fig 135; 813pp; English.	Qy	665	CGAGTTCCTAGCAGACGGTGTCTGCAAGAGGGGGACATAGCTCCCTCGGAGGGA	724
XX					
PS	AAS21244-AAS21518 encode for novel human secretory and transmembrane PRO	Db	947	CCAGTTCCTGGCAGACGATGGTCCGGCAAGAAAGGAGAGACACAGCTGCCCTGGGAGGGA	1006
CC	polypeptides. The PRO polypeptides are useful to detect other PRO	Qy	725	AGAAATCCAAAGAGAAAGCGCAGCGGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA	778
CC	polypeptides, to link bioactive molecules to cells expressing PRO	Db	1007	AGAGTCCAAAGAGAAAGCAGCAGGCGGCCAACAGGCGCAGGAGTAGCAGCAGCA	1066
CC	polypeptides, to modulate biological activities of cells expressing PRO				
CC	polypeptides, and to detect the presence of mammalian lung, colon,	Qy	779	AGCAGAGAAAGAACTGGGGGGCTGGGGGAGGATGCCAACCCGAGAGGAGGAGGTG	838
CC	breast, prostate, rectal, cervical or liver tumours by comparing PRO	Db	1067	AACAAAGGAAAGGAGCTGGGTGGCTTTGAGGGAGACCCCGCAGGAGGATGAGGGCA	1126
CC	polypeptide expression in a cell sample to that in a control sample. Some	Qy	839	TGCAGAGGAGATCGCCCTCCACACAGCCCTGATGAGCTGTGAGCCCGAGCTTAGTG	898
CC	of the 275 sequences are also useful to stimulate the release of tumour	Db	1127	TCCAGAGGAGATCCCTCTCTACACACAGCCCTGATGAGCTGTGAGCCCGAGCTTAGTG	1186
CC	necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or	Qy	899	TCCCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA	934
CC	differentiation of chondrocytes, the proliferation or gene expression in	Db	1187	CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA	1222
CC	pericyte cells, the release of proteoglycans from cartilage, the				
CC	proliferation of inner ear utricular supporting cells or of T-				
CC	lymphocytes, the release of a cytokine from peripheral blood monocytes				
CC	(PBMCS), or the proliferation of endothelial cells. Some of the PRO				
CC	polypeptides may modulate glucose or free fatty acid uptake by skeletal				
CC	muscle cells or by adipocytes; or inhibit binding of A-peptide to factor				
CC	VIIA. The PRO polypeptides can be used in assays to identify molecules				
CC	involved in binding interactions. The polynucleotides encoding PRO				
CC	polypeptides can be used to generate probes, antisense RNA/DNA,				
CC	transgenic or knock out animals and can be used in gene therapy				
XX					
SQ	Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;				
Query Match 37.6%; Score 691.6; DB 4; Length 1675;					
Best Local Similarity 84.5%; Pred. No. 7e-168;					
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;					
Qy	5 CGAGGAGAGACCGCGCGGTCCGCTCTGCTGGTCCGGCTGGGCCATGAGTCCA	64			
Db	287 CGCGGAGGAGAACCGCGCGTCTTTAGGTCCGGCCCGCGGCCCATGATTCAA	346			
Qy	65 TGTCTGAGTCCGCGCCCGCTCTTTATTTCTTCTGCTGCTGCTCCGCTGTGC	124			
Db	347 TGCCTGAGCCCGGTCCCGCTCTTCTGCTTCTTCCCTTGTCTGCTGCTGCTGC	406			
Qy	125 TCCTTCCTCCCGAAGTAGCCCGAGTCCCGCCGGGGCTGAGGAGACCGACTGGGTGC	184			
Db	407 TGTCTCCGCGCCCGGAGCTGGGGCCCGAGCCAGGCGGAGCTGAGGAGAACGACTGGGTTC	466			
Qy	195 GATTGCCAGCAATGCCAGTGTGCAAGTATGTTGCTGAGCTGAAGTCGGCTTTTG	244			
Db	467 GCCTGCCAGCAAAATGCCAAGTGTGTAATATGTTGCTGAGCTGAAGTCAGCCTTTG	526			
Qy	245 AGGAAACGGGAAAGAACCAAGAAAGTGTGACACCGGCTATGGCATCTGGACGGGAAGG	304			
Db	527 AGGAAACGGGAAAGAACCAAGAAAGTGTGATGACCGGCTATGGCATCTGGACGGGAAGG	586			
Qy	305 GCTCTGGAGTCAAGTACACCAAGTCCGACTTACCGGTTAATTGAAGTCACTGAGACCAATT	364			
Db	587 CCTCTGGAGTCAAAATACCAAGTCCGACTTGGCGTTAATGAAAGTCACTGAGACCAATT	646			
Qy	365 GCAGAGGCTTCTGCACTACGCTGACAGGAGGAGCTGGCAGCAACCGTTTGCA	424			
Db	647 GCAAGAGGCTCTCTGATTTATACCTTGCACAGGAGGAGCAGGCAATCGATTGCA	706			
Qy	425 AGGCTATGTCGAGAGACCTTTGAGACGCTGCACAACTAGTGCACAAAGGGGTCAAGGTGG	484			

RESULT 7
ACA03670
ID ACA03670 standard; cDNA; 1675 BP.
XX ACA03670;
AC ACA03670;
XX
DT 23-MAY-2003 (first entry)
XX
DE cDNA encoding human PRO polypeptide #68.
XX
KW Human; PRO polypeptide; secreted and transmembrane protein;
KW tumour necrosis factor-alpha; TNF-alpha; blood; proliferation;
KW differentiation; chondrocyte; tumour; genetic disorder; cytostatic; gene;
ss.
XX
OS Homo sapiens.
XX
PN US2003036180-A1.
XX
PD 20-FEB-2003.
XX
PF 09-MAY-2002; 2002US-00143114.
XX
PR 31-MAR-1997; 97WO-US005230.
PR 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US014552.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.
PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022591.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025108.

[illegible]

01-MAR-2001; 2001WO-US006666.
PR 09-MAR-2001; 2001US-00802706.
PR 14-MAR-2001; 2001US-00808689.
PR 22-MAR-2001; 2001US-00816744.
PR 05-APR-2001; 2001US-00828366.
PR 10-MAY-2001; 2001US-00854208.
PR 10-MAY-2001; 2001US-00854280.
PR 18-MAY-2001; 2001US-00860216.
PR 25-MAY-2001; 2001US-00866028.
PR 25-MAY-2001; 2001US-00866034.
PR 25-MAY-2001; 2001WO-US017092.
PR 01-JUN-2001; 2001US-00872035.
PR 01-JUN-2001; 2001WO-US017800.
PR 05-JUN-2001; 2001US-00874503.
PR 14-JUN-2001; 2001US-00882636.
PR 19-JUN-2001; 2001US-00886342.
PR 20-JUN-2001; 2001WO-US019692.
PR 21-JUN-2001; 2001US-00887879.
PR 22-JUN-2001; 2001WO-US020116.
PR 29-JUN-2001; 2001WO-US021066.
PR 09-JUL-2001; 2001WO-US021735.
PR 18-JUL-2001; 2001US-00908827.
PR 06-AUG-2001; 2001US-00924419.
PR 09-AUG-2001; 2001US-00927796.
PR 16-AUG-2001; 2001US-00931836.
PR 19-DEC-2001; 2001US-00028072.
XX
XX
(GETH) GENENTECH INC.
XX
XX
Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
PI Grittens ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2003-148238/14.
DR P-PSDB; ABU59718.
XX
XX Two hundred and seventy five nucleic acids encoding PRO polypeptides,
PT useful for treating pericyte-associated tumors, diabetes and various bone
PT and/or cartilage disorders, e.g. arthritis.
XX
XX Claim 2; Fig 135; 659pp; English.
XX
XX The invention describes an isolated human PRO polypeptide. The PRO
CC polypeptides are useful in detecting PRO polypeptides in a sample, in
CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and
CC in modulating at least one biological activity of a cell expressing a PRO
CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus
CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186
CC stimulate adrenal cortical capillary endothelial growth, and PRO536,
CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO19, PRO1126,
CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus
CC useful for treating conditions or disorders where angiogenesis would be
CC beneficial, e.g. wound healing and antagonist of this polypeptide are
CC useful for treating cancerous tumors. PRO812 inhibits vascular
CC endothelial growth factor (VEGF) stimulated proliferation of endothelial
CC cells and is thus useful for inhibiting endothelial cell growth in
CC mammals which would be beneficial in inhibiting tumour growth. PRO826,
CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of
CC stimulated T-lymphocytes and are therapeutically useful for enhancing
CC immune response. PRO828, PRO1068 or PRO1132 enhance survival of
CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of
CC rod photoreceptor cells) and therefore are useful for treating retinal
CC disorders of injuries, e.g. retinitis pigmentosa, AMD. PRO819, PRO813
CC and PRO11066 induce proliferation of mammalian kidney mesangial cells,
CC and therefore are useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Berger disease or other
CC nephropathies associated with dermatitis, herpetiformis or Crohn's
CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the
CC proliferation and/or redifferentiation of chondrocytes in culture and are
CC thus useful for treating sports injuries, and arthritis. This sequence
CC encodes a novel human PRO protein
XX
XX Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;
SQ

specification, or of the DNA deposited under any of the American Type Culture Collection (ATCC) Accession Numbers listed in the specification. Also included are a vector comprising the novel nucleic acid, a host cell comprising the vector, producing a PRO polypeptide, the isolated PRO polypeptides detailed above, a chimaeric molecule comprising the PRO polypeptide of fused to a heterologous amino acid sequence, an anti-PRO antibody, detecting a PRO polypeptide in a sample suspected of containing the PRO polypeptide, linking a bioactive molecule to a cell expressing a PRO polypeptide, modulating at least one biological activity of a cell expressing a PRO polypeptide, stimulating the release of tumour necrosis factor-alpha (TNF-alpha) from human blood, (or proteoglycans from cartilage or cytokine from peripheral blood mononuclear cells (PBMC)), modulating the uptake of glucose or FFA by skeletal muscle cells or adipocyte cells, stimulating the proliferation or differentiation of chondrocyte cells (or proliferation of or gene expression in pericyte cells), stimulating the proliferation of inner ear utricular supporting cells (or of T-lymphocyte cells, or of endothelial cells), inhibiting the binding of A-peptide to factor VIIA, or differentiation of adipocyte cells, detecting the presence of a tumour in a mammal and an oligonucleotide probe derived from any of the nucleotide sequences given in the specification. The polynucleotide is useful in molecular biology, including uses as hybridisation probes, in chromosome and gene mapping, in generating antisense RNA and DNA, and in gene therapy. The polynucleotide may also be used in preparing PRO polypeptides by recombinant techniques, and in generating either transgenic animals or knock-out animals which, in turn, are useful in the development and screening of therapeutically useful reagents. The PRO polypeptide or the antibody is used in preparing a medicament for treating a condition responsive to the polypeptide or antibody, such as tumours, and in various diagnostic assays. The present sequence encodes a PRO polypeptide

Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;

Qy	545	AGCAGTGTACGTCGCTGGTGGAGAGCTTTCAAGAGGTGATTGAGGACTGGTACAGGACC	604
Db	827	AGCAGTGTGATGTGCTGGTGGAGAGCTTTGAGGAGGTGATCGAGGACTGGTACAGGACC	886
Qy	605	ACCAGGAGGAAGACCTGACTGAATTCTCTGTGCCCAACCAACGTCGTGAAAGGGAAGGACA	664
Db	887	ACCAGGAGGAAGACCTGACTGAATTCTCTGTGCCCAACCAACGTCGTGAAAGGGAAGGACA	946
Qy	665	CGAGTTTCCTTAGCAGAGCGGTGCTCTGGCAAGAGGGGGGCATAGCCTCCCTCGGAGGGGA	724
Db	947	CCAGTTTCCTTGGCAGAGCAGTCGTCTCCGGCAAGAGGGGAGACACAGCTGCCTCGGAGGGGA	1006
Qy	725	AGAAATCCAAAGAAAGACCGACGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA	778
Db	1007	AGAAGTCCAAAGAAAGAGAGCAGCAGGCGCCAAAGCAGCAGCGCGCAGAGTAGCAGCAGCA	1066
Qy	779	AGCAGAGGAAGGAACCTGGGGGGCGCTGGGGGAGGATGCCAAACGCCGAGGAGGAGGAGGTG	838
Db	1067	AACAAGGAAGGAGCTGGGTGGCTTTGAGGGAGACCCACGCCCGAGGAGGATGAGGGCA	1126
Qy	839	TGCAGAAGGCATCGCCCTCCCAACACAGCCCCCTCATGAGCTGTGAGCCCCAGCTTAGTG	898
Db	1127	TCCAGAAGGCATCCCTCTCACACACAGCCCCCTCATGAGCTGTGAGCCCCAGCTTAGTG	1186
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGA	934
Db	1187	CCTCTGTCTGAGACCCCTGATTTTGAAGCTGAGGA	1222
RESULT 10			
ACA04091			
ID ACA04091 standard; cDNA; 1675 BP.			
XX	ACA04091;		
XX	AC AC XX		
XX	AC AC XX		
DT	27-MAY-2003 (first entry)		
XX	Human cDNA encoding a secreted/transmembrane protein, SEQ ID 135.		
XX	Human; ss; gene; secreted protein; transmembrane protein; PRO;		
KW	inflammatory disease; organ failure; atherosclerosis; cardiac injury;		
KW	infertility; birth defects; premature aging; AIDS; biosensor;		
KW	acquired immunodeficiency syndrome; cancer; diabetic complication;		
KW	bioreactor; tumour.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	US2003032155-A1.		
PN	13-FEB-2003.		
XX	03-MAY-2002; 2002US-00137865.		
PD	31-MAR-1997; 97WO-US005230.		
PF	12-JUN-1998; 98WO-US012456.		
XX	14-JUL-1998; 98WO-US014552.		
PR	28-AUG-1998; 98WO-US017888.		
PR	10-SEP-1998; 98WO-US018824.		
PR	14-SEP-1998; 98WO-US019093.		
PR	14-SEP-1998; 98WO-US019094.		
PR	14-SEP-1998; 98WO-US019177.		
PR	16-SEP-1998; 98WO-US019330.		
PR	17-SEP-1998; 98WO-US019437.		
PR	07-OCT-1998; 98WO-US021141.		
PR	29-OCT-1998; 98WO-US022991.		
PR	20-NOV-1998; 98WO-US024855.		
PR	01-DEC-1998; 98WO-US025108.		
PR	05-JAN-1999; 99WO-US000106.		
PR	08-MAR-1999; 99WO-US005028.		
PR	10-MAR-1999; 99WO-US005190.		
PR	20-APR-1999; 99WO-US008615.		
PR	14-MAY-1999; 99WO-US010733.		

CC	affinity purification of PRO from recombinant cell culture or natural
CC	sources. (I) and (II) are useful for tissue typing. This sequence encodes
CC	a novel human secreted and transmembrane PRO polypeptide.
XX	
SQ	Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;
	Query Match 37.6%; Score 691.6; DB 9; Length 1675;
	Best Local Similarity 84.5%; Pred. No. 7e-168;
	Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
Qy	5 CGAGGGAGGAAGCCGCCCGGCTCTGCTCTGGTTCGGCTGGGCGCATGAGTCCA 64
Db	287 CGCGGAGGAGGAACCCCGCGTCTTTAGGGTCCGGCCCGCGCCATGGAATCAA 346
Qy	65 TGTCTGAGCTCGCGCCCGCTGCTCTTATTTCTTTGCTGCTCTCTCCGCTGCTGC 124
Db	347 TGCCTGAGCCCGGCTCCCGTCTTCTGCTTCTTCTTCTGCTGCTGCTGCTGCTGC 406
Qy	125 TCCTTCTCTGCCCGGAAGCTAGGCCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184
Db	407 TGTCTCGCGCCCGGAGCTGGCCCGAGCCAGGCCGCGAGCTGAGGAGAACGACTGGGTTC 466
Qy	185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTG 244
Db	467 GCCTGCCCAGCAATGCGAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 526
Qy	245 AGGAACGGGAAGAACCAAGGAGTGAATGACACCGGCTATGGCATCTCGACCGGAAGG 304
Db	527 AGGAACCGGCAAGCAAGAGGAGTGTGGCAGCGGCTATGGCATCTCGACCGGAAGG 586
Qy	305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTTACGGTTAAATTTGAAGTCACTGAGACCAATTT 364
Db	587 CCTCTGGAGTCAATACACCAAGTCGGACTTTGCGTTAAATTCGAAGTCACTGAGACCAATTT 646
Qy	365 GCAAGAGCTTCTGGACTACAGCTGCAACAAGAGAGAGACTGGCAGCAACCGGTTTGGCA 424
Db	647 GCAAGAGCTTCTGGATTATAGCTTACAGCTGCAACAAGAGAGAGACCGGCAAGCAATCGATTTGGCA 706
Qy	425 AGGTATGTGGAGACCTTTGAGACGCTGCACACCTAGTCCACAAAGGGGTCAAGGTGG 484
Db	707 AGGCATGTGAGACCTTTGAGACATTTACAAACCTGGTACAAAGGGGTCAAGGTGG 766
Qy	485 TGATGATATCCCTTATGAGCTGTGGAAACGAGACCTTCAGCAGAGGTGGCTGACCTCAAGA 544
Db	767 TGATGACATCCCTTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA 826
Qy	545 AGCAGTGTGAGCTGCTGGTGGAAAGAGTTTGAAGAGGTGATTTGAGAGCTGTGTACAGAACCC 604
Db	827 AGCAGTGTGATGTCTGTTGGAAGAGTTTTCAGGAGGTGATTCGAGGACTGTGTACAGGAACC 886
Qy	605 ACCAGGAGGAGACCTGACTGAAATTCCTGTGTCCAAACCTGCTGAAGGGAAGGAGCA 664
Db	887 ACCAGGAGGAGACCTGACTGAAATTCCTGTGTCCAAACCTGCTGAGGAGGGAAGAGCA 946
Qy	665 CGAGTGTCTAGCAGAGCGGTGCTGTGGCAAGAGGGGACATAGCCTTCCCTGGAGGGA 724
Db	947 CCAGTTGCTTGGCAGAGCAGTGGTCCGCAAGAGGGAGACACAGCTGCCCTGGGAGGA 1006
Qy	725 AGAATCCAGAAAGAACGCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db	1007 AGAAGTCCAGAAAGAGACAGCAGGCGCAAGGCGGAGGAGTGTAGCAGCAGCA 1066
Qy	779 AGCAGAGGAGGAGTGGGGGCTGGGGAGGATGCCAACGCCGAGGAGGAGGAGGTG 838
Db	1067 AACAAAGAGGAGTGGGTGGCTTGGGGAGACCCCGCCCGAGGAGGATGAGGGA 1126
Qy	839 TGCAGAGGAGTCCGCCCTCCACACAGACCCCTCTGATGAGCTGTGTAGCCCGAGCTTAGTG 898
Db	1127 TCCAGAGGAGTCCCTCTTCCACACAGACCCCTCTGATGAGCTGTGTAGCCCGAGCTTAGTG 934
Qy	899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db	1187 CCTCTGCTGAGAGCCCTCTGATTTTGAAGCTGAGGA 1222

10-NOV-2000; 2000WO-US030873.
01-DEC-2000; 2000WO-US032678.
20-DEC-2000; 2000US-00747259.
20-DEC-2000; 2000WO-US034956.
28-FEB-2001; 2001US-00796498.
28-FEB-2001; 2001WO-US006520.
01-MAR-2001; 2001WO-US006666.
09-MAR-2001; 2001US-00802706.
14-MAR-2001; 2001US-00808689.
22-MAR-2001; 2001US-00816744.
05-APR-2001; 2001US-00828366.
10-MAY-2001; 2001US-00854208.
10-MAY-2001; 2001US-00854280.
18-MAY-2001; 2001US-00860216.
25-MAY-2001; 2001US-00866028.
25-MAY-2001; 2001US-00866034.
25-MAY-2001; 2001WO-US017092.
01-JUN-2001; 2001US-00872035.
01-JUN-2001; 2001WO-US017800.
05-JUN-2001; 2001US-00874503.
14-JUN-2001; 2001US-00882636.
19-JUN-2001; 2001US-00886342.
20-JUN-2001; 2001WO-US019692.
21-JUN-2001; 2001US-00887879.
22-JUN-2001; 2001WO-US020116.
29-JUN-2001; 2001WO-US021066.
09-JUL-2001; 2001WO-US021735.
18-JUL-2001; 2001US-00908827.
06-AUG-2001; 2001US-00924419.
09-AUG-2001; 2001US-00927796.
16-AUG-2001; 2001US-00931836.
19-DEC-2001; 2001US-00028072.
(GETH) GENENTECH INC.
Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
WPI; 2003-584997/55.
P-PSDB; ADA45655.
Novel secreted and transmembrane polypeptide for modulating biological
activity of cell expressing the polypeptide, identifying agonists or
antagonists of polypeptide, and as molecular weight markers.
Claim 2; Fig 135; 659pp; English.
The invention describes 305 nucleic acids encoding PRO (secreted and
transmembrane) polypeptides (I). (I) is useful for stimulating the
release of TNF-alpha from human blood, for modulating the uptake of
glucose or FFA by skeletal muscle cells or adipocyte cells, for
stimulating the proliferation or differentiation of chondrocyte cells,
for stimulating the proliferation of or gene expression in pericyte
cells, for stimulating the release of proteoglycans from cartilage, for
stimulating the proliferation of inner ear utricular supporting cells,
for stimulating the proliferation of T-lymphocyte cells, for stimulating
the release of a cytokine from PBMC cells, for inhibiting the binding of
A-peptide to factor VIIa, for inhibiting the differentiation of adipocyte
cells, for stimulating proliferation of endothelial cells, for detecting
the presence of tumour in a mammal. The tumour is lung, colon, breast,
prostate, rectal, cervical or liver tumour. The oligonucleotide probes
are useful for isolating genomic and cDNA nucleotide sequences or
antisense probes. (I) is also useful as therapeutic agent. PRO is useful
in assays to identify other proteins or molecules involved in binding
interaction. A polynucleotide (II) encoding (I) is useful in chromosome
and gene mapping, in generation of antisense RNA and DNA, in the
preparation of PRO polypeptide, for generating transgenic animals or
knockout animals which in turn are useful in the development and
screening of therapeutically useful reagents, in gene therapy, for
chromosome identification, as chromosome marker, and for generating
probes. An anti-(I)-antibody is useful in diagnostic assays for PRO, e.g.
detecting its expression in specific cells, tissues or serum, and for

RESULT 12

ADA76085
ID ADA76085 standard; cDNA; 1675 BP.

AC ADA76085;

XX 20-NOV-2003 (first entry)

XX Human PRO polynucleotide #68.

XX Human; gene; ss; PRO; secreted polypeptide; transmembrane polypeptide;
KW tumour necrosis factor-alpha; TNF-alpha; chondrocyte cell; tumour;
KW cancer; adrenal; lung; colon; breast; prostate; rectum; kidney; cervix;
KW liver; microvascular endothelial cell; glucose; FFA;
KW skeletal muscle cell; adipocyte cell; pericyte cell;
KW inner ear utricular supporting cell; T-lymphocyte cell;
KW endothelial cell tube formation; bone disorder; cartilage disorder;
KW sports injury; proteoglycan; articular cartilage defect; osteoarthritis;
KW rheumatoid arthritis; haemoglobin-associated disorder thalassaemia;
KW immune system cell infiltration.

XX Homo sapiens.

XX US2003073212-A1.

XX 17-APR-2003.

XX 16-APR-2002; 2002US-00123903.

XX 31-MAR-1997; 97WO-US005230.
PR 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US014552.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.
PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 14-SEP-1998; 98WO-US019177.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022991.
PR 29-OCT-1998; 98WO-US022992.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025108.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 22-DEC-1999; 99WO-US030720.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000277.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005746.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 30-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.
PR 10-NOV-2000; 2000WO-US030873.
PR 01-DEC-2000; 2000WO-US032678.
PR 20-DEC-2000; 2000US-00747259.
PR 20-DEC-2000; 2000WO-US034956.
PR 28-FEB-2001; 2001US-00796498.
PR 28-FEB-2001; 2001WO-US006520.
PR 01-MAR-2001; 2001WO-US006666.
PR 09-MAR-2001; 2001US-00802706.
PR 14-MAR-2001; 2001US-00808689.
PR 22-MAR-2001; 2001US-00816744.
PR 05-APR-2001; 2001US-00828366.
PR 10-MAY-2001; 2001US-00854208.
PR 10-MAY-2001; 2001US-00854280.
PR 18-MAY-2001; 2001US-00860216.
PR 25-MAY-2001; 2001US-00866028.
PR 25-MAY-2001; 2001US-00866034.
PR 25-MAY-2001; 2001WO-US017092.
PR 01-JUN-2001; 2001US-00872035.
PR 01-JUN-2001; 2001WO-US017800.
PR 05-JUN-2001; 2001US-00874503.
PR 14-JUN-2001; 2001US-00882636.
PR 19-JUN-2001; 2001US-00886342.
PR 20-JUN-2001; 2001WO-US019892.
PR 21-JUN-2001; 2001US-00887879.
PR 22-JUN-2001; 2001WO-US020116.
PR 29-JUN-2001; 2001WO-US021066.
PR 09-JUL-2001; 2001WO-US021735.
PR 18-JUL-2001; 2001US-00908827.
PR 06-AUG-2001; 2001US-00924419.
PR 09-AUG-2001; 2001US-00927796.
PR 16-AUG-2001; 2001US-00931836.
PR 19-DEC-2001; 2001US-00028072.

(GETH) GENENTECH INC.

Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

WPI; 2003-687639/65.
P-PSDB; ADA76086.

XX New isolated nucleic acid encoding a secreted and transmembrane
PT polypeptide, designated e.g. PRO1114 or PRO4978, useful in chromosome and
PT gene mapping, in generating antisense RNA and DNA, and in gene therapy.
XX Claim 2; Fig 135; 659pp; English.
XX


```
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 22-DEC-1999; 99WO-US030720.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005746.
PR 02-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 30-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.
PR 10-NOV-2000; 2000WO-US030873.
PR 01-DEC-2000; 2000WO-US032678.
PR 20-DEC-2000; 2000US-00747259.
PR 20-DEC-2000; 2000WO-US034956.
PR 28-FEB-2001; 2001US-00796498.
PR 28-FEB-2001; 2001WO-US006520.
PR 01-MAR-2001; 2001WO-US006666.
PR 09-MAR-2001; 2001US-00802706.
PR 14-MAR-2001; 2001US-00808689.
PR 22-MAR-2001; 2001US-00816744.
PR 05-APR-2001; 2001US-00828366.
PR 10-MAY-2001; 2001US-00854208.
PR 18-MAY-2001; 2001US-00854280.
PR 25-MAY-2001; 2001US-00860216.
PR 25-MAY-2001; 2001US-00866034.
PR 25-MAY-2001; 2001WO-US017092.
PR 01-JUN-2001; 2001US-00872035.
PR 01-JUN-2001; 2001WO-US017800.
PR 05-JUN-2001; 2001US-00874503.
PR 14-JUN-2001; 2001US-00886236.
PR 19-JUN-2001; 2001US-00886342.

PR 20-JUN-2001; 2001WO-US019692.
PR 21-JUN-2001; 2001US-00887879.
PR 22-JUN-2001; 2001WO-US020116.
PR 29-JUN-2001; 2001WO-US021066.
PR 09-JUL-2001; 2001WO-US021735.
PR 18-JUL-2001; 2001US-00908827.
PR 06-AUG-2001; 2001US-00924419.
PR 09-AUG-2001; 2001US-00927796.
PR 16-AUG-2001; 2001US-00931836.
PR 19-DEC-2001; 2001US-00028072.
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2003-521854/49.
DR P-PSDB; ADA18736.
XX
XX New PRO nucleic acid, useful for preparing a composition for treating
e.g., tumors.
XX
XX Claim 2; Fig 135; 660pp; English.
XX
XX The invention relates to isolated human PRO polypeptides (secreted and
transmembrane polypeptides) and the polynucleotides encoding them. The
invention also relates to an antibody which specifically binds to a PRO
polypeptide, a method for stimulating the release of tumour necrosis
factor-alpha (TNF-alpha) from human blood, a method for stimulating the
proliferation or differentiation of chondrocyte cells and a method for
detecting the presence of a tumour in a mammal (e.g. lung, colon, breast,
prostate, rectal, cervical and liver tumours). The polynucleotides are
useful in molecular biology, including uses as hybridisation probes, in
chromosome and gene mapping, in generating antisense RNA and DNA and in
gene therapy. The polynucleotides may also be used in preparing PRO
polypeptides by recombinant techniques and in generating either
transgenic animals or knock-out animals which are useful in the
development and screening of therapeutically useful reagents. The PRO
polypeptides or antibodies are used in preparing a medicament for
treating a condition responsive to the polypeptides or antibodies, such
as tumours, for modulating the uptake of glucose or FFA by adipocyte
cells, for stimulating the proliferation of or gene expression in
pericyte cells, for stimulating the release of proteoglycans from
cartilage, for stimulating the proliferation of inner ear utricular
supporting cells, for stimulating the release of cytokines from PBMC
cells, for inhibiting the binding of A-peptide to factor VIIA, for
inhibiting the differentiation of adipocyte cells and for stimulating the
proliferation of endothelial cells. This sequence represents a human PRO
polynucleotide of the invention. Note: The sequence data for this patent
is also available in electronic format from USPTO at
seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 U; 0 Other;

Query Match 37.6%; Score 691.6; DB 9; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7e-168;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY 5 CGAGGAGGAGCGCCGCGCTCTCTCTGCGTCCGCTCGGCTCGGCTGGGCCATGGAGTCCA 54
DB 287 CGCGGAGGAGGAGAACCCGCGCTCTCTTAGGGTCCGGCCCGCGCGGCGCATGGATTCAA 346
QY 65 TGTCTGAGCTCGCGCCCGCTGCTCTATTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 124
DB 347 TGCCTGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 406
QY 125 TCCTTCCTGCCCCGAAGCTAGCGCCCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGC 184
DB 407 TGCTGCGGGCCCCGAGCTGGGCCCCGAGCCAGCCAGCGGAGCTGAGGAGAACGACTGGGTTC 466
QY 185 GATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGTCTTTTG 244
```

Db	467	GCCTGCCCGCCAGCAATGCGAAGTGTGTAAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG	526	OS	and.		
Qy	245	AGGAAACGGGAAAGACCAAGGAAAGTGAATTCACACCGGCTATGGCATCTGCGACGGGAAGG	304	OS	transmembrane.		
Db	527	AGGAAACGGGCAAGACCAAGAGGTGATTGGCAGGGCTATGCGCATCTGGACCGAAGG	586	OS	protein.		
Qy	305	GCTCTGGAGTCAAGTACACCAAGTCGAGCTTACCGTTTAATTGAAGTCACTGAGACCATTT	364	OS	PRO4409.		
Db	587	CCTCTGGAGTCAAATACCAAGTCGGACTTGGCGTTTAATCGAAGTCACTGAGACCATTT	646	OS	cDNA.		
Qy	365	GCAAGAGGCTTCTGGACTACAGCTGCACAAAGGAGGAGCTGGCAGCAACCGGTTTGCCA	424	XX	US2003049816-A1.		
Db	647	GCAAGAGGCTCCTGGATTATAGCTGTCACAAAGGAGAGGACCGGAGCAATTCGATTTGCCA	706	XX	13-MAR-2003.		
Qy	425	AGGTTATGTCGGAGACCTTTGAGAGCTGCACAACTAGTCCACAAAGGGGTCAAAGTGG	484	XX	15-APR-2002; 2002US-00123262.		
Db	707	AGGCATGTCAGAGACCTTTGAGACATTTACAACTGGTACAAAGGGGTCAAAGTGG	766	XX	31-MAR-1997; 97WO-US005230.		
Qy	485	TGATGGATATCCCTCATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA	544	XX	12-JUN-1998; 98WO-US012456.		
Db	767	TGATGGACATCCCTCATGAGCTGTGGAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA	826	XX	14-JUL-1998; 98WO-US014552.		
Qy	545	AGCAGTGTGAGTGTCTGGTGAAGAGTTTGAAGAGTGATTGAGGACTGTGTACAGGAACC	604	XX	28-AUG-1998; 98WO-US017888.		
Db	827	AGCAGTGTGATGTCTGTGTGAAGAGTTTGAAGAGGTGATCGAGGACTGGTACAGGAACC	886	XX	10-SEP-1998; 98WO-US018824.		
Qy	605	ACGAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACCTGCTGAAGGGGAAAGGACA	664	XX	14-SEP-1998; 98WO-US019094.		
Db	887	ACGAGGAGGAGACCTGACTGAATTCCTCTGTGCCAACACCTGCTGAGGGGAAAGGACA	946	XX	16-SEP-1998; 98WO-US019177.		
Qy	665	CGAGTTGCCCTAGCAGAGCGGTGGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGGA	724	XX	17-SEP-1998; 98WO-US019437.		
Db	947	CCAGTTGCCCTGGCAGAGCAGTGGTCCGGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA	1006	XX	07-OCT-1998; 98WO-US021141.		
Qy	725	AGAAATCCAAAGAAAGCGGACGGAGTCAAG-----GGCTCCTCAGTGGCGAGCAGCA	778	XX	29-OCT-1998; 98WO-US022991.		
Db	1007	AGAAATCCAAAGAAAGCAGCAGCAGGCGCCAAAGCGCAGCAGCGGAGAGTAGCAGCAGCA	1066	XX	20-NOV-1998; 98WO-US024855.		
Qy	779	AGCAGAGGAGGAACCTGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGAGGGTG	838	XX	01-DEC-1998; 98WO-US025108.		
Db	1067	AACAAGGAAGGAGCTGGGTGGCTTTGAGGAGACCCCGAGGAGGAGGAGGAGTGAAGGGA	1126	XX	05-JAN-1999; 99WO-US000106.		
Qy	839	TGCAGAGAGGCATCCGCCCTCCACACACAGCCCCCTGATGAGTGTGTGAGCCCGAGCTTAGTG	898	XX	08-MAR-1999; 99WO-US005028.		
Db	1127	TCCAGAGGCGATCCCTCTCACACACAGCCCCCTGATGAGTCTGAGCCCAACCCAGCAT	1186	XX	10-MAR-1999; 99WO-US005190.		
Qy	899	TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGA 934		XX	14-MAY-1999; 99WO-US010733.		
Db	1187	CCTCTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 1222		XX	02-JUN-1999; 99WO-US012252.		
RESULT 14					XX	01-SEP-1999; 99WO-US020111.	
ID	ADA61358 standard; cDNA; 1675 BP.					XX	08-SEP-1999; 99WO-US020594.
XX						XX	13-SEP-1999; 99WO-US020944.
AC	ADA61358;					XX	15-SEP-1999; 99WO-US021090.
XX						XX	15-SEP-1999; 99WO-US021547.
DT	20-NOV-2003 (first entry)					XX	05-OCT-1999; 99WO-US023089.
XX						XX	29-NOV-1999; 99WO-US028214.
DE	Homo sapiens.					XX	30-NOV-1999; 99WO-US028313.
XX						XX	01-DEC-1999; 99WO-US028301.
KW	Human; secreted and transmembrane protein; PRO; gene; ss;					XX	01-DEC-1999; 99WO-US028634.
KW	tumour necrosis factor alpha release; TNF-alpha release;					XX	02-DEC-1999; 99WO-US028551.
KW	glucose uptake modulator; FFA uptake modulator;					XX	02-DEC-1999; 99WO-US028564.
KW	cell proliferation stimulator; cell differentiation stimulator;					XX	02-DEC-1999; 99WO-US028565.
KW	cell differentiation inhibitor; cytokine release stimulator; tumour;					XX	16-DEC-1999; 99WO-US030095.
KW	lung tumour; colon tumour; breast tumour; prostate tumour; rectal					XX	20-DEC-1999; 99WO-US030911.
KW	tumour; cervical tumour; liver tumour; chromosome mapping; gene					XX	22-DEC-1999; 99WO-US030999.
KW	therapy; chromosome identification; chromosome marker.					XX	30-DEC-1999; 99WO-US0310720.
XX						XX	30-DEC-1999; 99WO-US031243.
OS	Novel.					XX	30-DEC-1999; 99WO-US031274.
OS	human.					XX	05-JAN-2000; 2000WO-US000219.
OS	secreted.					XX	06-JAN-2000; 2000WO-US000277.
OS						XX	11-FEB-2000; 2000WO-US000376.
						XX	18-FEB-2000; 2000WO-US003565.
						XX	18-FEB-2000; 2000WO-US004341.
						XX	22-FEB-2000; 2000WO-US004342.
						XX	24-FEB-2000; 2000WO-US004414.
						XX	24-FEB-2000; 2000WO-US004914.
						XX	24-FEB-2000; 2000WO-US005004.
						XX	01-MAR-2000; 2000WO-US00505004.
						XX	02-MAR-2000; 2000WO-US005601.
						XX	02-MAR-2000; 2000WO-US005746.
						XX	02-MAR-2000; 2000WO-US005841.
						XX	10-MAR-2000; 2000WO-US006319.
						XX	15-MAR-2000; 2000WO-US006684.
						XX	20-MAR-2000; 2000WO-US007377.
						XX	21-MAR-2000; 2000WO-US007532.
						XX	30-MAR-2000; 2000WO-US008439.
						XX	17-MAY-2000; 2000WO-US013705.
						XX	22-MAY-2000; 2000WO-US014042.
						XX	30-MAY-2000; 2000WO-US014941.

Db 1127 TCCAGAGGCATCCCTCTCACACAGCCCCCTGATGAGCTCTGAGCCACCAGCAT 1186

Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934

Db 1187 CCTCTGCTCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

RESULT 15

ADB19143

ID ADB19143 standard; cDNA; 1675 BP.

XX

AC ADB19143;

XX

DT 20-NOV-2003 (first entry)

XX

XX Novel human secreted and transmembrane protein PRO4409 cDNA.

XX

XX Human; secreted and transmembrane protein; PRO; gene; ss;

KW Tumour necrosis factor alpha release; TNF-alpha release;

KW glucose uptake modulator; FFA uptake modulator;

KW cell proliferation stimulator; cell differentiation stimulator;

KW cell differentiation inhibitor; cytokin.

XX

OS Homo sapiens.

XX

XX US2003068796-A1.

XX

PD 10-APR-2003.

XX

PF 15-APR-2002; 2002US-00123261.

XX

PR 31-MAR-1997; 97WO-US005230.

PR 12-JUN-1998; 98WO-US012456.

PR 14-JUL-1998; 98WO-US014552.

PR 28-AUG-1998; 98WO-US017888.

PR 10-SEP-1998; 98WO-US018824.

PR 14-SEP-1998; 98WO-US019031.

PR 14-SEP-1998; 98WO-US019034.

PR 14-SEP-1998; 98WO-US019177.

PR 16-SEP-1998; 98WO-US019330.

PR 17-SEP-1998; 98WO-US019437.

PR 07-OCT-1998; 98WO-US021141.

PR 29-OCT-1998; 98WO-US022991.

PR 29-OCT-1998; 98WO-US022992.

PR 20-NOV-1998; 98WO-US024855.

PR 01-DEC-1998; 98WO-US025108.

PR 05-JAN-1999; 99WO-US000106.

PR 08-MAR-1999; 99WO-US005028.

PR 10-MAR-1999; 99WO-US005190.

PR 20-APR-1999; 99WO-US008615.

PR 14-MAY-1999; 99WO-US010733.

PR 02-JUN-1999; 99WO-US012252.

PR 01-SEP-1999; 99WO-US020111.

PR 08-SEP-1999; 99WO-US020594.

PR 13-SEP-1999; 99WO-US020944.

PR 15-SEP-1999; 99WO-US021090.

PR 15-SEP-1999; 99WO-US021547.

PR 05-OCT-1999; 99WO-US023089.

PR 29-NOV-1999; 99WO-US028214.

PR 30-NOV-1999; 99WO-US028313.

PR 30-NOV-1999; 99WO-US028409.

PR 01-DEC-1999; 99WO-US028301.

PR 01-DEC-1999; 99WO-US028634.

PR 02-DEC-1999; 99WO-US028551.

PR 02-DEC-1999; 99WO-US028564.

PR 02-DEC-1999; 99WO-US028565.

PR 16-DEC-1999; 99WO-US030095.

PR 20-DEC-1999; 99WO-US030911.

PR 20-DEC-1999; 99WO-US030999.

PR 22-DEC-1999; 99WO-US030720.

PR 30-DEC-1999; 99WO-US031243.

PR 30-DEC-1999; 99WO-US031274.

PR 05-JAN-2000; 2000WO-US000219.

PR 06-JAN-2000; 2000WO-US000277.

PR 06-JAN-2000; 2000WO-US000376.

PR 11-FEB-2000; 2000WO-US003565.

PR 18-FEB-2000; 2000WO-US004341.

PR 18-FEB-2000; 2000WO-US004342.

PR 22-FEB-2000; 2000WO-US004414.

PR 24-FEB-2000; 2000WO-US004914.

PR 24-FEB-2000; 2000WO-US005004.

PR 01-MAR-2000; 2000WO-US005501.

PR 02-MAR-2000; 2000WO-US005746.

PR 02-MAR-2000; 2000WO-US005841.

PR 10-MAR-2000; 2000WO-US006319.

PR 15-MAR-2000; 2000WO-US006884.

PR 20-MAR-2000; 2000WO-US007377.

PR 21-MAR-2000; 2000WO-US007532.

PR 30-MAR-2000; 2000WO-US008439.

PR 17-MAY-2000; 2000WO-US013705.

PR 22-MAY-2000; 2000WO-US014042.

PR 30-MAY-2000; 2000WO-US014941.

PR 02-JUN-2000; 2000WO-US015264.

PR 28-JUL-2000; 2000WO-US020710.

PR 11-AUG-2000; 2000WO-US022031.

PR 23-AUG-2000; 2000WO-US023522.

PR 24-AUG-2000; 2000WO-US023328.

PR 08-NOV-2000; 2000WO-US030952.

PR 10-NOV-2000; 2000WO-US030873.

PR 01-DEC-2000; 2000WO-US032578.

PR 20-DEC-2000; 2000US-00747259.

PR 20-DEC-2000; 2000WO-US034956.

PR 28-FEB-2001; 2001US-00796498.

PR 28-FEB-2001; 2001WO-US006520.

PR 01-MAR-2001; 2001WO-US006666.

PR 09-MAR-2001; 2001US-00802706.

PR 14-MAR-2001; 2001US-00808689.

PR 22-MAR-2001; 2001US-00816744.

PR 05-APR-2001; 2001US-00828366.

PR 10-MAY-2001; 2001US-00854208.

PR 10-MAY-2001; 2001US-00854280.

PR 18-MAY-2001; 2001US-00860216.

PR 25-MAY-2001; 2001US-00866028.

PR 25-MAY-2001; 2001US-00866034.

PR 25-MAY-2001; 2001WO-US017092.

PR 01-JUN-2001; 2001US-00872035.

PR 01-JUN-2001; 2001WO-US017800.

PR 05-JUN-2001; 2001US-00874503.

PR 14-JUN-2001; 2001US-00882836.

PR 19-JUN-2001; 2001US-00886342.

PR 20-JUN-2001; 2001WO-US019692.

PR 21-JUN-2001; 2001US-00887879.

PR 22-JUN-2001; 2001WO-US020116.

PR 29-JUN-2001; 2001WO-US021066.

PR 09-JUL-2001; 2001WO-US021735.

PR 18-JUL-2001; 2001US-00908827.

PR 06-AUG-2001; 2001US-00924419.

PR 09-AUG-2001; 2001US-00927796.

PR 16-AUG-2001; 2001US-00931836.

PR 19-DEC-2001; 2001US-00028072.

XX

FA (GETH) GENENTECH INC.

XX

XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX

DR WPI; 2003-695927/66.

DR P-PSDB; ADB19144.

XX

XX Novel secreted and transmembrane PRO polypeptides useful for stimulating

PT the release of tumor necrosis factor alpha and detecting the presence of

PT a tumor in a mammal.

XX

PS Claim 2; Fig 135; 660pp; English.

XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:40:35 ; Search time 355 Seconds
(without alignments)
9208.270 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcc.....taaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/1_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467	25.4	1201	3	US-09-023-655-140
2	458.6	24.9	1534	3	Sequence 140, App
3	160.2	8.7	744	3	Sequence 274, App
4	160.2	8.7	1512	3	Sequence 3, Appli
5	156.6	8.5	1658	3	Sequence 1, Appli
6	75.2	4.1	7218	2	Sequence 23, Appl
7	59.4	3.2	24207	2	Sequence 14, Appl
8	56.2	3.1	7218	2	Sequence 14364, A
9	55.4	3.0	1926	3	Sequence 14, Appl
10	55.4	3.0	1931	2	Sequence 4, Appli
11	49.4	2.7	3130	3	Sequence 2, Appli
12	49.4	2.7	87734	3	Sequence 5779, Ap
13	48.6	2.6	3489	2	Sequence 17521, A
14	48.6	2.6	3489	2	Sequence 1, Appli
15	48.6	2.6	3489	3	Sequence 1, Appli
16	48.6	2.6	3489	3	Sequence 1, Appli
17	48.6	2.6	32207	2	Sequence 20, Appl
18	48.6	2.6	32207	3	Sequence 20, Appl
19	48.6	2.6	32207	3	Sequence 20, Appl
20	47.4	2.6	601	3	Sequence 91307, A
21	47.4	2.6	12695	3	Sequence 16775, A
22	46.8	2.5	289	3	Sequence 17, Appl
23	46.8	2.5	289	3	Sequence 17, Appl
24	46.8	2.5	15252	3	Sequence 13584, A

25	46.4	2.5	64309	3	US-09-949-016-14581	Sequence 14581, A
26	46.4	2.5	767677	3	US-09-949-016-12147	Sequence 12147, A
27	46.4	2.5	767677	3	US-09-949-016-17361	Sequence 17361, A
28	46	2.5	18955	3	US-09-949-016-13343	Sequence 13343, A
29	46	2.5	30678	3	US-09-949-016-12818	Sequence 12818, A
30	45	2.4	474	3	US-09-621-976-18033	Sequence 18033, A
31	44.8	2.4	3053	3	US-10-104-047-958	Sequence 958, App
32	44.6	2.4	3319	2	US-08-006-676B-2	Sequence 2, Appli
33	44.6	2.4	3319	2	US-08-282-845-1	Sequence 1, Appli
34	44.6	2.4	3319	2	US-08-428-414A-4	Sequence 4, Appli
35	44.6	2.4	3319	6	PCT-US94-00324-2	Sequence 2, Appli
36	44.4	2.4	1926	3	US-09-249-585A-2	Sequence 2, Appli
37	44.4	2.4	1926	3	US-09-410-339-3	Sequence 3, Appli
38	44.4	2.4	2580	3	US-09-050-863-2	Sequence 2, Appli
39	44.4	2.4	2580	3	US-09-359-081-2	Sequence 2, Appli
40	44.4	2.4	5452	2	US-09-130-114-1	Sequence 1, Appli
41	44.4	2.4	8705	3	US-09-647-344A-14	Sequence 14, Appl
42	44.4	2.4	9600	3	US-08-910-647-1	Sequence 1, Appli
43	44.4	2.4	9600	3	US-09-620-925-1	Sequence 1, Appli
44	44.4	2.4	10596	2	US-07-884-811-15	Sequence 15, Appl
45	44.4	2.4	10596	2	US-07-885-971-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1

US-09-023-655-140
; Sequence 140, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THPINOB01
; CLONE: 030137

US-09-023-655-140

```
Query Match      25.4%; Score 467; DB 3; Length 1201;
Best Local Similarity 79.0%; Pred. No. 4.1e-109;
Matches 612; Conservative 0; Mismatches 95; Indels 68; Gaps 2;

QY 228 GCTGAAGTCGGCTTTTGAGGAAACGGGAAAGACCAAGGAAGTGATTTGACACCGGCTATGG 287
Db 1 GCTGAGTCAGCCTTTTGAGGAAACCGGCAAGCCAGGAGGTGATTTGGCACGGGCTATGG 60

QY 288 CATCTGGAACGGGAGGCTCTGGAGTCAAGTACACCAAGTC----- 329
Db 61 CATCTGGAACGAGGCTCTGGAGTCAAAATACAAAGTCCATTTTCAGATCCCCCAGA 120

QY 330 -----GGACTTACGTTAATT 345
Db 121 CCAGATGACCTATCTTCTTCAGCTCTGAGTCACTTCCCAATGGGACTTCGGGTAAATC 180

QY 346 GAAGTCACTGAGACCATTTTGAAGAGGCTTCTGGACTACAGCTGCACAAAGGAGGAGT 405
Db 181 GAAGTCACTGAGACCATTTTGAAGAGGCTTCTGGATTAAGCTGCACAAAGGAGGAGT 240

QY 406 GCGAGCAACCGGTTTCCAAAGGTTATGTCGAGACCTTTTGAGACGCTTGCAACACTAGTC 465
Db 241 GCGAGCAATCGATTTGCCAAGGGCATGTCAGAGACCTTTGAGACATTAACAACCTGGTA 300

QY 466 CACAAGGGGTCAAGGTGTTGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCA 525
Db 301 CACAAGGGGTCAAGGTGTTGATGATATCCCTATGAGCTGTGGAACGAGACCTTCGCA 360

QY 526 GAGGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGGTGGAGAGTGTGAAGAGGTGATT 585
Db 361 GAGGTGGCTGACCTCAAGAAGCAGTGTGATGTGCTGGTGGAGAGTGTGAAGAGGTGATC 420

QY 586 GAGGACTGTTACAGGAACACCAAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACAC 645
Db 421 GAGGACTGTTACAGGAACACCAAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACAC 480

QY 646 GTGCTGAAGGGAAGGACACAGATTGCTTAGCAGACGGTGGTCTGGCAAGAGGGGAC 705
Db 481 GTGCTGAAGGGAAGGACACCAAGTTGCTTGGCAGAGCAGTGGTCCGGCAAGAGGGAGAC 540

QY 706 ATAGCTTCCCTGGGAGGGAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAG-----GGC 759
Db 541 ACAGCTGCCCTGGGAGGGAAGAGTCCAAAGAAAGACAGCAGAGGCCCAAGGCAGCAGGC 600

QY 760 TCCTCCAGTGGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCCTGGGGGAGGATGCCAAC 819
Db 601 GGCAGGATGAGCAGCAGCAAAACAAAGGAAGGAGCTGGGTGGCTTTGAGGGGAGACCCACG 660

QY 820 GCGAGGAGGAGGAGGAGTGGCAGAGGATCGCCCTCCACACAGCCCCCTGTATGAG 879
Db 661 CCCGAGGAGGATGAGGGGATCCAGAAGGCATCCCTCTCTCACACACAGCCCCCTGTATGAG 720

QY 880 CTGTGAGCCCAAGCTTAGTGTCTTCAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 721 CTCTGAGCCCAAGCAGCATCTCTGTCTGTGAGACCCCTGATTTGAAGCTGAGGA 775
```

RESULT 2

```
US-09-023-655-274
; Sequence 274, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
```

```
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA: US/09/023,655
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 274:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT07
; CLONE: 1298861
; US-09-023-655-274
```

```
Query Match      24.9%; Score 458.6; DB 3; Length 1534;
Best Local Similarity 84.8%; Pred. No. 6.3e-107;
Matches 525; Conservative 0; Mismatches 92; Indels 2; Gaps 2;

QY 5 CGAGGGAGGAAGCGCCCGGGTCCGCTCTGCTCTGGTCCGGTGGGCCATGGAGTCCA 64
Db 660 CGCGGGAGGAGGAACCCCGGGTCTTTAGGGTCCGGGCCCGGGCCATGGAATCAA 719

QY 65 TGTCTGAGCTCGCGCCCGGCTCTTATTTCTTTGCTGCTGCTGCTTC-CGCTGCTG 123
Db 720 TGCTTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTTGTCTGCTGATGANNNNNN 779

QY 124 CTCCTTCTCTCCCGAAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTG 183
Db 780 NNNNNNNNGCCCGGAGCTGGGCCCGAGCCGAGCTGAGGAGAACGACTGGGT 839

QY 184 CGATTGCCAGCAATCGGAAGTGTGCAAGTATGTTGCTGTGAGCTG-AAAGTCGGCTTT 242
Db 840 CGCTGCGCCAGCAATCGAAGTGTGTAATATGTTGCTGTGAGCTGTAAAGTCAGCCTT 899

QY 243 TGAGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTTGGACGGGAA 302
Db 900 TGAGAAACCGGCAAGCAAGGAGGTGATTTGGCACGGGCTATGGCATCTTGGACCGAA 959

QY 303 GGGCTCTGGAGTCAAGTACACCAAGTCGAGTTCGAGTTCGAGTTCAGTTCAGTGCAT 362
Db 960 GGCCTCTGGAGTCAATATACACCAAGTCGAGTTCGAGTTCGAGTTCAGTTCAGTGCAT 1019

QY 363 TTGCAAGAGGCTTCTGAGCTACGCTGCACAAAGGAGGACTGGCAGCAACCGGTTGTC 422
Db 1020 TTGCAAGAGGCTTCTGAGTATATAGCTTGCACAAAGGAGGACCGGCGAGCAATCGAT 1079

QY 423 CAAGGATGTCGAGAGACCTTTTGAGAGCTGCACAACTAGTCCACAAAGGGGTCAAGGT 482
Db 1080 CAAGGATGTCGAGAGACCTTTTGAGAGCTGCACAACTAGTCCACAAAGGGGTCAAGGT 1139

QY 483 GGTGATGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGTGCACCTCAA 542
Db 483 GGTGATGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGTGCACCTCAA 542
```

Db 1140 GGTGATGACATCCCTATGAGCTGTGGAACGAGACTTCTGACAGGTGGCTGACCTCAA 1199
Qy 543 GAAGCAGTGTGACGTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGTACAGAA 602
Db 1200 GAAGCAGTGTGATGTGCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGTGTACAGAA 1259
Qy 603 CCACCAGGAGGAGACCTG 621
Db 1260 CCACCAGGAGGAGACCTG 1278

RESULT 3

US-09-163-285-3
; Sequence 3, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MMI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..744
US-09-163-285-3

Query Match 8.7%; Score 160.2; DB 3; Length 744;
Best Local Similarity 57.3%; Pred. No. 6.9e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
Qy 162 GCCTCAGAGACCGACTGGTCCGATTGCCAGCAATGCCAAGTGTGCAAGTATGTTGC 221
Db 75 GGAGGAGACGATGACACAGAACGCTTCCCGAGCAATGCCAGTGTGTAGCTGCTGAG 134
Qy 222 TGTGAGCTGAAGTCGGCTTTTGAGGAACCGGAAAGCAAGGAGTGTGACACCGG 281
Db 135 CACAGACTACAGCGGGAACCTGAGTCGCACCGGTGCATCTCAGAGGTGCTGGAGCTGG 194
Qy 282 CTATGGCATCTGGACGGGAAGGGCTCTGGAGTCAAG---TACACCAAGTCGGACTTACG 338
Db 195 GCAGGTGCTGGATACAGGCAAGAGAGAGACACGTGCTTACAGCGTTTCAGAGACAAG 254

Qy 339 GTTAATTGAAGTCACTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCTGCACAGGA 398
Db 255 GCTGGAAGAGCGCTTAGAGAAATTTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA 314
Qy 399 GAGGACTGGCAGCAACCGTTTGGCAAGGGTATGTCCGAGAGCCTTTTGAGAGCCTGCACAA 458
Db 315 CGCAAGGGCTCACTAGATATGCCAAGGGTCAGAGTCAGACCATGGCAACACTGAAGG 374
Qy 459 CCTAGTCCAAAGGGGTCAAGGGTGTGATGATATCCCTATGAGCTGTGGAACGAGAC 518
Db 375 CCTAGTGCAGAAAGGGGTGAAGGTGGATCTCGGGATCCCTCTGGAGCTTTGGGATGAGCC 434
Qy 519 CTCAGCAGAGTGGCTGACCTCAAGAACCATGTGACGTGCTGGTGGAGAGTTTGAAGA 578
Db 435 CAGCGTGGAGGTCACTACCTCAAGAACCATGTGAGACCATGTGTTGGAGGAGTTTGAAGA 494
Qy 579 GGTGATTGAGGACTGGTACAGAAACCCACGAGGAGAACCTGACTGAATTCCTCTGTGC 638
Db 495 CATTGTGGGAGACTGGTACTTCCACCATCAGAGAGCGCCCTACAAAATTTCTCTGTGA 554
Qy 639 CAACCACTGCTGAAGGGAAAGGACACGAGTTGCTAGCAGAGCGGTGCTGGCAAGAA 698
Db 555 AGTCACTGTCTCCAGCTGCTGAAACTGCATGCTACAGGAAACTTGGACTGGAAAGGA 614
Qy 699 G 699
Db 615 G 615

RESULT 4

US-09-163-285-1
; Sequence 1, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MMI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: 125...868
US-09-163-285-1

Query Match      8.7%; Score 160.2; DB 3; Length 1512;
Best Local Similarity 57.3%; Pred. No. 9e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 162 GGCTGAGGAGACCGACTGGTGGATTCGCCAGCAAAATGCCAAGTGTGCAAGTATGTTGC 221
Db 199 GGAGGAGGACGATGATACAGACAAGCGCTTGGCCAGCAAAATGCCAAGTGTGTAAGTGCCTGAG 258
Qy 222 TGTGAGCTGAAGTGGCTTTTGGAGAAACGGGAAAGACCAAGGAAGTGAATGACACCGG 281
Db 259 CACAGAGCTACAGGGGAACTGAGTCGCACCGTGCATCTCGAGAGTGTCTGGAGCTGGG 318
Qy 282 CTATGGCATCTCGACGGGAAGGGCTCTGGAGTCAAG---TACCAAGTCCGACTTACG 338
Db 319 GCAGGTGCTGGATACAGGCAAGAGAGAGACACGTCGCTTTACAGCGTTCAGAGACAAG 378
Qy 339 GTTAATTCAGTCACTGAGACCAATTGCAAGAGGCTTCTGCACTACAGCCTGCACAGGA 398
Db 379 GCTGGAAGAGCCTTAGAGAAATTTATGTGAGCGGATCCTGGACTATAGTGTTCACGCTGA 438
Qy 399 GAGGACTGGCAGCAACCGGTTTGGCAAGGGTATGTCCGAGACCTTTGAGACGCTGCACAA 458
Db 439 CGCAAGGGCTCACTGAGATATGCCAAGGGTCAAGTCAAGCATGGCAACACTGAAAGG 498
Qy 459 CCTAGTCACAAAAGGGGTCAAGGTGGTGATGGATATCCCTATGAGCTGTGGAAACGAGAC 518
Db 499 CCTAGTCAGAAAGGGGTGAAGTGGATCTCGGGATCCCTCTGGAGCTTTGGGATGAGCC 558
Qy 519 CTCAGCAGAGTGGCTGACCTCAAGAAGCAGTGTGACGTGCTGGTGGGAAGTGTGAAGA 578
Db 559 CAGCGTGGAGGTGCATACATACCTCAAGAAGCAGTGTGAGACCATGTTTGGAGGAGTTGAAGA 618
Qy 579 GGTGATTGAGACTGGTACAGAACCCAGGAGGAAGACCTGACTGAATTCCTCTGTGC 638
Db 619 CATTGGGAGACTGGTACTTCCACCATCAGAGCAGCCCTTACAAATTTCTCTGTGA 678
Qy 639 CAACCACGTGCTGAAGGGAAGGACACGAGTTGGCTAGCAGCGGTGGTCTGGCAAGAA 698
Db 679 AGGTCACTGCTCCCGAGCTGCTGAACTGCATGCTCAGAGGAACCTTGGACTGGAAAGGA 738
Qy 699 G 699
Db 739 G 739

RESULT 5
US-09-482-273-23/c
; Sequence 23, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030PL
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-482-273-23
Query Match      8.5%; Score 156.6; DB 3; Length 1658;
Best Local Similarity 76.2%; Pred. No. 7.8e-30;
Matches 208; Conservative 0; Mismatches 59; Indels 6; Gaps 1;

Qy 668 GTTCCCTTAGCAGACCGGTGGTCTGGCAAGAAGGGGGACATAGCCCTCCCTGGGAGGGAAGA 727
Db 308 GTTCCCTGGCAGACAGTGGTCCGGCAAGAAGGGAGACACAGACTGCCCTGGGGGAAGA 249
Qy 728 AATCAAGAAGAACCGCAGCGGAGTCAAG-----GGTCTCTCCAGTGGGAGCAGCAGCAGC 781
Db 248 AGTCCAAGAAGAACGAGCAGCGGCGCAAGGCGCAGCGGCGGAGGAGTAGCAGCAGCAAAAC 189
Qy 782 AGAGGAAGGAACTGGGGGGCTGGGGAGGATGCCAAGCGCGAGGAGGAGGAGGAGGTGTGC 841
Db 188 AAAGGAAGGAGCTGGGTGGGCTTGGGAGACCCAGCCCGGAGGAGATGAGGCGATCC 129
Qy 842 AGAAGGATCGCCCTCCACACAGCCCTGATGAGCTGTGAGCCCGAGCTTAGTGTGCC 901
Db 128 AGAAGGATCCCCCTCTCACACAGCCCTCCCTGATGAGCTCTGAGCCCGAGCAGCATCCT 69
Qy 902 TTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 68 CTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 36

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
```

[illegible]

; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5779
 ; LENGTH: 3130
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-5779

Query Match 2.7%; Score 49.4; DB 3; Length 3130;
 Best Local Similarity 47.3%; Pred. No. 0.023;
 Matches 181; Conservative 0; Mismatches 201; Indels 1; Gaps 1;
 QY 189 GCCAGCAAAATGCGAAGTGTGCA-AGTATGTTGCTGTGGAGTGAAGTCGGCTTTTGAGG 247
 DB 1745 GCGGGCCCAACGAGAAAGTACGACAGGAGGTGGCGGGCTTGAAGGACAAGGTTTCAGCAGG 1804
 QY 248 AAACGGGAAAGCAACGAGAAAGTGTGACACCGGCTATGGCATCTCGACGGGAAGGGCT 307
 DB 1805 CCACAGCGAGAACATGGGGCTAATGGAACTGGAATCCAGCTGACTCGCTGGGCT 1864
 QY 308 CTGGAGTCAAGTACACCAAGTCGGAATTTACGGTTAAATTGAAGTCACTGAGACCAATTTGCA 367
 DB 1865 CGGACCACCAAGAGTCCCTGGAGGACCTCAAGCCACCTGAACTCGGGCCCGAGCGGCC 1924
 QY 368 AGAGGCTTCTGCACTACAGCTGCAACAGGAGGAGTGGCAGCAACCGGTTTTCGAAGG 427
 DB 1925 AGCAGAAAGAGATCGCGGAGCTGGAAGGAGTGAAGGAGTGAAGGAGGATCAAGATGGAGCACCAGC 1984
 QY 428 GTATGTCGAGACCTTTCAGAGCTGCAACACCTAGTCCAAAGGGGTCAAGTGTGA 487
 DB 1985 TGGAGCTGGGTAACCTTTCAGGCGCAAGCATGACCTGGAGACCGCATGACCTGAAGGAGA 2044
 QY 488 TGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGC 547
 DB 2045 AGGAGGCTTCGAGAGAAAGTGCAGGAGGCGCCAGGAGGAGTGGCTGGCTGCAGCGGC 2104
 QY 548 AGTGTGAGCTGCTGTGGAAGAG 570
 DB 2105 ACTGGCGGGCCAGCTGGAGGTG 2127

RESULT 12
 US-09-949-016-17521
 ; Sequence 17521, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT FILING DATE: 2000-04-14
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 17521
 ; LENGTH: 87734
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-17521

Query Match 2.7%; Score 49.4; DB 3; Length 87734;

Best Local Similarity 47.3%; Pred. No. 0.082;
 Matches 181; Conservative 0; Mismatches 201; Indels 1; Gaps 1;
 QY 189 GCCAGCAAAATGCGAAGTGTGCA-AGTATGTTGCTGTGGAGTGAAGTCGGCTTTTGAGG 247
 DB 60303 GCGGGCCCAACGAGAAAGTACGACAGGAGGTGGCGGGCTTGAAGGACAAGGTTTCAGCAGG 60362
 QY 248 AAACGGGAAAGCAACGAGAAAGTGTGACACCGGCTATGGCATCTCGACGGGAAGGGCT 307
 DB 60363 CCACAGCGAGAACATGGGGCTAATGGAACTGGAATCCAAAGCTGCACTCGCTGGGCT 60422
 QY 308 CTGGAGTCAAGTACACCAAGTCGGAATTTACGGTTAAATTGAAGTCACTGAGACCAATTTGCA 367
 DB 60423 CGGACCACCAAGAGTCCCTGGAGGACCTCAAGCCACCTGAACTCGGGCCCGAGCGGCC 60482
 QY 368 AGAGGCTTCTGCACTACAGCTGCAACAGGAGAGGAGTGGCAGCAACCGGTTTTCGAAGG 427
 DB 60483 AGCAGAAAGAGATCGCGGAGCTGGAAGGAGTGAAGGAGGATCAAGATGAGCACCAGC 60542
 QY 428 GTATGTCGAGACCTTTCAGAGCTGCAACACCTAGTCCAAAGGGGTCAAGTGTGA 487
 DB 60543 TGGAGCTGGGTAACCTTTCAGGCGCAAGCATGACCTGGAGACCGCATGACCTGAAGGAGA 60602
 QY 488 TGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGC 547
 DB 60603 AGGAGGCTTCGAGAGAAAGTGCAGGAGGCGCCAGGAGGAGTGGCTGGCTGCAGCGGC 60662
 QY 548 AGTGTGAGCTGCTGTGGAAGAG 570
 DB 60663 ACTGGCGGGCCAGCTGGAGGTG 60685

RESULT 13
 US-08-728-323A-1
 ; Sequence 1, Application US/08728323A
 ; Patent No. 5948676
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
 ; Sarcoma-Associated Herpesvirus, DNA
 ; TITLE OF INVENTION: Encoding Same And Uses Thereof
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,323A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-278-0400
 ; TELEFAX: 212-391-0525
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3489 base pairs
 ; TYPE: nucleic acid

[illegible]

Db 2600 AGGAGCAAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAG 2640

Search completed: December 30, 2005, 03:44:24
Job time : 359 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 22:55:25 ; Search time 1381 Seconds
(without alignments)
11011.865 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413489005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691.6	37.6	1378	6	US-10-126-103-139 Sequence 139, App
2	691.6	37.6	1378	7	US-10-431-096-139 Sequence 139, App
3	691.6	37.6	1675	5	US-10-028-072-135 Sequence 135, App
4	691.6	37.6	1675	5	US-10-140-808-135 Sequence 135, App
5	691.6	37.6	1675	5	US-10-121-049-135 Sequence 135, App
6	691.6	37.6	1675	5	US-10-123-904-135 Sequence 135, App
7	691.6	37.6	1675	5	US-10-140-470-135 Sequence 135, App
8	691.6	37.6	1675	5	US-10-175-746-135 Sequence 135, App
9	691.6	37.6	1675	5	US-10-176-918-135 Sequence 135, App
10	691.6	37.6	1675	5	US-10-176-921-135 Sequence 135, App
11	691.6	37.6	1675	5	US-10-137-865-135 Sequence 135, App
12	691.6	37.6	1675	5	US-10-140-474-135 Sequence 135, App
13	691.6	37.6	1675	5	US-10-142-431-135 Sequence 135, App
14	691.6	37.6	1675	5	US-10-143-114-135 Sequence 135, App
15	691.6	37.6	1675	5	US-10-142-419-135 Sequence 135, App
16	691.6	37.6	1675	5	US-10-123-262-135 Sequence 135, App
17	691.6	37.6	1675	5	US-10-142-423-135 Sequence 135, App
18	691.6	37.6	1675	5	US-10-121-050-135 Sequence 135, App
19	691.6	37.6	1675	5	US-10-141-755-135 Sequence 135, App
20	691.6	37.6	1675	5	US-10-143-032-135 Sequence 135, App
21	691.6	37.6	1675	5	US-10-123-108-135 Sequence 135, App
22	691.6	37.6	1675	5	US-10-123-236-135 Sequence 135, App
23	691.6	37.6	1675	5	US-10-123-261-135 Sequence 135, App

ALIGNMENTS

RESULT 1

US-10-126-103-139
; Sequence 139, Application US/10126103
; Publication No. US20030224486A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB PATHW
; FILE REFERENCE: D0108.np
; CURRENT APPLICATION NUMBER: US/10/126.103
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/284,962
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/286,645
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/346,986
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 139
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-126-103-139

Query Match	37.6%	Score 691.6;	DB 6;	Length 1378;
Best Local Similarity	84.5%	Pred. No. 6.6e-197;		
Matches 791;	Conservative 0;	Mismatches 139;	Indels 6;	Gaps 1;

QY	5	CGAGGAGGAAGCCCGCGGTCTCTCTCTGGTCCGCTGGGCCATGGAGTCA	64
DB	21	CGCGGAGGAGAAACCGCCCGTCTTTAGGTCGGGCGCGCCATGGATTCAA	80
QY	65	TGCTGAGCTCGCCCGCTCTTTATTTCTTCTGCTGCTGCTGCTGCTGCTG	124
DB	81	TGCTGAGCCCGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	140
QY	125	TCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	184
DB	141	TGCTGCGCCCGCGAGCTGGCCCGAGCTGGCCCGAGCTGGAGAACACTGGGTTT	200
QY	185	GATTCGCCAGCAATGCGAAGTGTGCAAGTATGTTCTGTGAGCTGAAGTCGGCTTTTG	244
DB	201	GCCTGCCAGCAATGCGAAGTGTGCAAGTATGTTCTGTGAGCTGAAGTCAGCCTTTG	260
QY	245	AGGAAACGGGAAACACCAAGGATGTTGACCGGCTATGGCATCTGGACCGGAGG	304

Db 261 AGGAAACCGGCAAGACCAAGGAGTGATTGGCAGCGGCTATGGCATCTGGCATCTGGACAGAAGG 320
Qy 305 GCTCTGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTT 364
Db 321 CCTCTGGAGTCAAAATACACCAAGTCGGACTTGGCGGTTAATCGAAGTCACTGAGACCAATTT 380
Qy 365 GCAAGAGGCTTCTGGACTACAGCCTGCGCAAGGAGGAGACTGGCAGCAACCGGTTTGGCA 424
Db 381 GCAAGAGGCTCTGGATTATAGCTGCGCAAGGAGGAGACCGGCAGCNAATCGATTTGGCA 440
Qy 425 AGGGTATGTCGGAGACCTTTGAGCGCTGCGCAACACTAGTCCACAAGAGGGGTCAAGGTGG 484
Db 441 AGGGCATGTACAGAGACCTTTGAGACATTACACAACCTGTACACAAGAGGGGTCAAGGTGG 500
Qy 485 TGATGGATATCCCTATAGCTGTGGAAAGGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 501 TGATGGACATCCCTATAGCTGTGGAAAGGAGACTTCTGCGAGAGGTGGCTGACCTCAAGA 560
Qy 545 AGCAGTGTGACGTGCTGGTGGAAAGGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 604
Db 561 AGCAGTGTGATGCTGGTGGAAAGGTTTGAAGAGGTGATTGAGGAGTGTATCGAGGACTGGTACAGGAACC 620
Qy 605 ACCAGGAGGAAGACCTGACTGAATTCCTCTGTGCGCAACCAACCTGCTGAAGGGGAAAGGACA 664
Db 621 ACCAGGAGGAAGACCTGACTGAATTCCTCTGTGCGCAACCAACCTGCTGAAGGGGAAAGGACA 680
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGCGCAAGAGGGGACATAGCCTCCCTGGAGGGGA 724
Db 681 CCAGTTGCTGCGCAGAGCAGTGGTCCGCGCAAGAGGGGACACAGCTGCTGGAGGGGA 740
Qy 725 AGAAATCCAAAGAGCGGCGAGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db 741 AGAAGTCCAAAGAGAGCAGCAGGCGCAAGGCGAGCGGCGGAGGTAGCAGCAGCA 800
Qy 779 AGCAGAGGAAGAACTGGGGGCGCTGGGGGAGGATGCCAACCCGCGAGGAGAGGGGTG 838
Db 801 AACAAAGGAAGGAGCTGGGTGGCTTTGAGGAGACCCCGAGCCCGGAGGAGATGAGGGCA 860
Qy 839 TGCAGAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 861 TCCAGAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 920
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 921 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 956

RESULT 2

US-10-431-096-139
; Sequence 139, Application US/104311096
; Publication No. US20040086896A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0108A CIP
; CURRENT APPLICATION NUMBER: US/10/431,096
; CURRENT FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: US 60/284,962
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 10/126,103
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/286,645
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/346,986
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 139
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-431-096-139

Query Match 37.6%; Score 691.6; DB 7; Length 1378;
Best Local Similarity 84.5%; Pred. No. 6.6e-197;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;
Qy 5 CGAGGAGGAGAGCCGCCGGGTCCGCTCTGCTCTGGGTCCGGTGGGCCCATGAGATCCA 64
Db 21 CGCGGAGGAGGAACCGCCGGTCTTTAGGGTCCGGGCCCGGCCCATGATTTCAA 80
Qy 65 TGTCTGAGCTCGCCCGCGCTGCTCTTATTTCTTTGCTGCTGCTGCTGCTGCTGCTG 124
Db 81 TGCCTGAGCCCGCTCCCGCTGCTCTGCTGCTTCTTCCCTTCTGCTGCTGCTGCTG 140
Qy 125 TCCTTCTGCTCCCGAAGCTAGGCCCGAGTCCCGCGGGCTGAGAGACCGACATGGGTGC 184
Db 141 TGCTGCCGCGCCCGAGCTGGGCCCGAGCCAGGCCGCGAGCTGAGAGAACGATCGGGTTC 200
Qy 185 GATTGCCCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTG 244
Db 201 GCCTGCCCCAGCAATGCGAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 260
Qy 245 AGGAAACCGGAAAGACAAGGAGTGTGACACCGGCTATGGCATCTTGACCGGAAAGG 304
Db 261 AGGAAACCGGCAAGACAAGGAGGTGATTGGCACCGGCTATGGCATCTTGACCGAAGG 320
Qy 305 GCTCTGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATTGAAGTCACTGAGACCAATTT 364
Db 321 CCTCTGAGTCAATACACCAAGTCGGACTTACGGTTAAATCGAAGTCACTGAGACCAATTT 380
Qy 365 GCAAGAGGCTTCTGGACTACAGCCTGCAAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 381 GCAAGAGGCTCTGGATTATAGCTGCAAGGAGAGGACCGGCGAGCAATCGATTTGCCA 440
Qy 425 AGGTATGTCGGAGACCTTTGAGACGCTGCAACACTAGTCCACAAGGGGTCAAGGTGG 484
Db 441 AGGCATGTGAGAGACCTTTGAGACATTACACAACCTGGTACACAAGGGGTCAAGGTGG 500
Qy 485 TGATGATATCCCTATGAGCTGTGGAAAGGAGCCTCAGCAGAGGTGGCTGACCTCAAGA 544
Db 501 TGATGAGACATCCCTATGAGCTGTGGAAAGGAGACTTCTGAGAGGTGGCTGACCTCAAGA 560
Qy 545 AGCAGTGTGAGTCTGCTGGTGGAAAGGTTTGAAGAGGTGATTTAGGACTGTGTTACAGGAACC 604
Db 561 AGCAGTGTGATGCTGCTGGTGGAAAGTTTGAGGAGGTGATCGAGGACTGGTACAGGAACC 620
Qy 605 ACCAGGAGGAAGACCTGACTGAATTTCTCTGTCGAACCAACCAGCTGTAAGGGAAGGACA 664
Db 621 ACCAGGAGGAAGACCTGACTGAATTTCTCTGCGCCAAACCAACGCTGTAAGGGGAAAGACA 680
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTGGGAGGGA 724
Db 681 CCAGTTGCTGCGCAGAGCAGTGGTCCGCAAGAGGGGAGACACAGCTGCTGGGAGGGA 740
Qy 725 AGAAATCCAAAGAAAGCGCAGCGGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db 741 AGAAGTCCAAAGAAAGAGCAGCAGGCGCAAGGCGAGCGGCGGAGGTAGCAGCAGCA 800
Qy 779 AGCAGAGGAAGAACTGGGGGCGCTGGGGGAGGATGCCAACCCGCGAGGAGAGGGGTG 838
Db 801 AACAAAGGAAGGAGCTGGGTGGCTTTGAGGAGACCCCGAGCCCGGAGGAGATGAGGGCA 860
Qy 839 TGCAGAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 861 TCCAGAGGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCCGAGCTTAGTG 920
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 921 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 956

RESULT 3

US-10-028-072-135
; Sequence 135, Application US/10028072

Publication No. US20030004311A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Laureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C182

; CURRENT APPLICATION NUMBER: US/10/140,808

; CURRENT FILING DATE: 2002-05-07

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 135

; LENGTH: 1675

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-140-808-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;

Best Local Similarity 84.5%; Pred. No. 7.1e-197;

Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

```
Qy 5 CGAGGAGGAGCGCGCGGTCGCTCTGCTCTGGGTCGGCTGGGCCATGAGTCCA 64
Db 287 CGCGGAGGAGGAAACCGCCCGGTCCTTTAGGGTCGGGCCCGCGGCCATGATCAA 346
Qy 65 TGTCTGAGCTCGCGCCCGCTCTCTTATTTCTTCTGCTGCTGCTTCCGCTGCTGC 124
Db 347 TGCCTGAGCCCGCTGCTGCTCTCTTCTTCTTCTTCTTCTGCTGCTGCTGCTGC 406
Qy 125 TCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
Db 407 TGTCTGCGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
Qy 185 GATTGCCAGCAATGCGAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 244
Db 467 GCCTGCCAGCAATGCGAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 526
Qy 245 AGGAAACGGGAAAGCAAGCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 304
Db 527 AGGAAACGGGAAAGCAAGCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 586
Qy 305 GCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364
Db 587 CCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 646
Qy 365 GCAAGAGGCTTCTGAGCTTACAGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 424
Db 647 GCAAGAGGCTTCTGAGCTTACAGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 706
Qy 425 AGGTTATCTCGAGACCTTTCGAGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 484
Db 707 AGGTTATCTCGAGACCTTTCGAGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGAGAGGTGCTGACCTCAAGA 544
Db 767 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGAGAGGTGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGTGCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 604
Db 827 AGCAGTGTGAGCTGTGCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 886
Qy 605 ACCAGGAGGAGACCTGACTGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 664
Db 887 ACCAGGAGGAGACCTGACTGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 946
Qy 665 CGAGTTGCTAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 724
Db 947 CGAGTTGCTAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1006
Qy 725 AGAAATCCAAGAGAGCGCGAGGCTCAAG-----GGTCTCTCCAGTGGCAGCAGA 778
Db 1007 AGAAGTCCAAGAGAGCGCGAGGCTCAAG-----GGTCTCTCCAGTGGCAGCAGA 1066
Qy 779 AGCAGAGAGGAACTGGGGGCTGCGGAGGATGCCAAAGCAGGAGGAGGAGGAGG 838
Db 1067 ACAAAGAGGAGGAGCTGGGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1126
Qy 839 TGCAGAGGCTCGCCCCCTCCACACAGCCCCCTGATGAGTGTGAGCCAGCTTAGTG 898
```

```
Db 1127 TCAGAGGAGATCCCTCTCTACACAGCCCCCTGATGAGCTTGAGCCCAACCCAGCAT 1186
Qy 899 TCCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA 934
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222
```

RESULT 5

US-10-121-049-135

; Sequence 135, Application US/10121049

; Publication No. US20030022239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C17

; CURRENT APPLICATION NUMBER: US/10/121,049

; CURRENT FILING DATE: 2002-04-12

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 135

; LENGTH: 1675

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-121-049-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;

Best Local Similarity 84.5%; Pred. No. 7.1e-197;

Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

```
Qy 5 CGAGGAGGAGCGCGCGGTCGCTCTGCTCTGGGTCGGCTGGGCCATGAGTCCA 64
Db 287 CGCGGAGGAGGAAACCGCCCGGTCCTTTAGGGTCGGGCCCGCGGCCATGATCAA 346
Qy 65 TGTCTGAGCTCGCGCCCGCTCTTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 124
Db 347 TGCCTGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 406
Qy 125 TCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184
Db 407 TGTCTGCGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
Qy 185 GATTGCCAGCAATGCGAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 244
Db 467 GCCTGCCAGCAATGCGAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 526
Qy 245 AGGAAACGGGAAAGCAAGCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 304
Db 527 AGGAAACGGGAAAGCAAGCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 586
Qy 305 GCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364
Db 587 CCTCTGAGTCAAGTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 646
Qy 365 GCAAGAGGCTTCTGAGCTTACAGCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 424
```

```
Db 647 GCAAGAGGCTCTCGGATTATAGCTTGCAACAGGAGAGGACCGCGCAGCAATCGATTTGGCA 706
Qy 425 AGGGTATCTCGAGACCTTTGAGACGCTGCACAACTAGTGCACAAAAGGGGTCAAGGTGG 484
Db 707 AGGGCATGTGAGAGACCTTTGAGACATTTACAACTGTTGTATACAAAGGGGTCAAGGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAACGAGACTTCTGCAGAGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGCTGTGGAAGAGTTTGAAGAGTGATTCAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGAGTGTGCTGTGGAAGAGTTTGAAGAGTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAAGACTGACTCAATTTCTCTGTGCCAACCACTGCTGAAGGGAAGAGACA 664
Db 887 ACCAGGAGGAAGACTGACTCAATTTCTCTGTGCCAACCACTGCTGAAGGGAAGAGACA 946
Qy 665 CGAGTTGCTTAGCAGAGCGGTGGTCTGGCAAGAAAGGGGACATAGCCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTTGCAGAGCAGTGGTCCGCGCAAGAGGGAGACACAGCTGCCCTGGGAGGGA 1006
Qy 725 AGAATCCAAGAAAGCGAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db 1007 AGAAGTCCAAGAAAGAGCAGCAGCGGCCCAAGGAGCAGCGGCGGAGGTAGCAGCAGCA 1066
Qy 779 AGCAGGAAGAACTGGGGGCGCTGGGGAGGATGCCAACCGCGAGGAGGAGGGGTG 838
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGGAGGAGACCCAGCCCGGAGGAGATGAGGGA 1126
Qy 839 TGCAGAGGCAATCGCCCTCCACACACAGCCCGCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 1127 TCCAGAGGCAATCCCTCTCACACAGCCCGCTGATGAGCTCTGAGCCCAACCCAGCAT 1186
```

RESULT 6

```
US-10-123-904-135
; Sequence 135, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: ACIDS ENCODING THE SAME
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 135
; LENGTH: 1675
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-904-135
```

```
Query Match 37.6%; Score 691.6; DB 5; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7.1e-197;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

Qy 5 CGAGGAGGAAAGCCGCCCGGGTCCGCTCTGCTCTGGGTCCGGCTAGGATCCA 64
Db 287 CGCGGAGGAGAAACCCCGCGTCTTTTAGGGTCCGGCCCGCGGCCCATGGATTCAA 346
Qy 65 TGTCTGAGCTCGCGCCCGCTCCTTATTTCTTTTCTGCTGCTGCTCGCTCGCTGC 124
Db 347 TGCCTGAGCCCGCTCCGCTGCTCTTCTGCTTCTTCCCTGCTGCTGCTGCTGCTGC 406
Qy 125 TCCTTCTGCCCCGAAGCTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGC 184
Db 407 TGCTGCCGCCCGGAGCTGSGCCCGAGCCAGGCGCGAGCTGAGGAGAACACTGGGTTC 466
Qy 185 GATTGCCCAGCAAAATGCAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCGGCTTTTG 244
Db 467 GCCTGCCCAGCAAAATGCAAGTGTGTAATATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG 526
Qy 245 AGGAAACGGGAAAGACCAAGAACTGATTCACACCGGCTATGGCATCTCGACCGGAGG 304
Db 527 AGGAAACCGGCAAGACCAAGGAGGTGATTGGCACGGGCTATGGCATCTCGACCCAGAGG 586
Qy 305 GCTCTGAGTCAAGTACACCAAGTCGGACTTACCGGTTAAATTGAAAGTCACTGAGACCATTT 364
Db 587 CCTCTGAGTCAANAATACCAAGTCGGACTTTCGGTTAAATCGAAGTCACTGAGACCATTT 646
Qy 365 GCAAGAGGCTTCTGGACTACAGCTGCAAGGAGAGAGACTGGCAGCAACCGGTTTGGCA 424
Db 647 GCAAGAGGCTCCTGGATTTATAGCCTGCACAAGGAGAGACCGGAGCAATCGATTTGGCA 706
Qy 425 AGGATATGTCGGAGACCTTTGAGACGCTGCACAACCTAGTCCACAAGGGGTCAAGGTGG 484
Db 707 AGGCAATGTCAGAGACCTTTGAGACATTTACAACTTGGTTACAAAGGGGTCAAGGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAAACGAGACTTCTGACAGAGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGCTGGTGGAAAGAGTTTGAAGAGTGAATTGAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGATGCTGCTGGTGGAAAGATTGAGGAGGTGATCGAGGACTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAGACCTGACTCAATTTCTCTGTCGAACCCAGCTGCTGAAGGGAAGAGACA 664
Db 887 ACCAGGAGGAGACCTGACTCAATTTCTCTGTCGCAACCCAGCTGCTGAAGGGAAGAGACA 946
Qy 665 CGAGTTGCTTAGCAGAGCGGTGGTCTGGCAAGAGGGGACATAGCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTTGGCAGAGCAGTGGTCCGCGCAAGAGGAGAGACACAGCTGCCCTGGGAGGGA 1006
Qy 725 AGAAATCCAAGAAAGACCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGCA 778
Db 1007 AGAAGTCCAAGAAAGAGCAGCAGCGGCCCAAGGAGCAGCGGCGGAGGTAGCAGCAGCA 1066
Qy 779 AGCAGAGGAGGAACCTGGGGGCGCTGGGGAGGATGCCAACCGCGAGGAGGAGGGGTG 838
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGAAGGAGACCCCGAGGAGGATGAGGGA 1126
Qy 839 TGCAGAGGCAATCCCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 1127 TCCAGAGGCAATCCCTCTCACACAGCCCGCTGATGAGCTCTGAGCCCAACCCAGCAT 1186
Qy 899 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGA 934
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222
```

RESULT 7

```
US-10-140-470-135
; Sequence 135, Application US/10140470
```



```

Db 347 TGCCTGAGCCCGCGTCCCGCTGTCTCTCTGCTTCTTCCCTGCTGCTGCTGCTGC 406
Qy 125 TCCTTCTCTGCCCCGAAGCTAGCCCGAGTCCCGCGCGGCTGAGAGACCGACTGGGTGC 184
Db 407 TGCTGCGCGCCCCGAGAGCTGGCGCCGAGCCAGCGCGAGCTGAGAGAACACTGGGTTC 466
Qy 185 GATTGCCCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db 467 GCCTGCCCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTG 526
Qy 245 AGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGACCGGAGG 304
Db 527 AGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGACCGAAGG 586
Qy 305 GCTCTGGAGTCAAGTACACCAAGTCGGAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 364
Db 587 CCTCTGGAGTCAAAATACACCAAGTCGGAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 646
Qy 365 GCAAGAGGCTTCTGAGCTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGAGTATAGCTGTGCAACAGGAGAGGACCGCGCAATCGATTGCCA 706
Qy 425 AGGTATGTGCGAGACCTTTGAGAGCTGTGCAACACTTAGTGTGCAACAAAGGGGTCAAGTGG 484
Db 707 AGGGCATGTGAGAGACCTTTGAGACATTTACACAACTTGTGTATACAAAGGGGTCAAGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACTCTGAGAGAGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAACGAGACTCTTGCAGAGGTGGCTGACCTCAAGA 826
Qy 545 GCAAGAGGCTTCTGAGCTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGAGTATAGCTGTGCAACAGGAGAGGACCGCGCAATCGATTGCCA 706
Qy 425 AGGTATGTGCGAGACCTTTGAGAGCTGTGCAACACTTAGTGTGCAACAAAGGGGTCAAGTGG 484
Db 707 AGGGCATGTGAGAGACCTTTGAGACATTTACACAACTTGTGTATACAAAGGGGTCAAGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACTCTGAGAGAGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAACGAGACTCTTGCAGAGGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGTGCTGTGGAAGAGTGTGGAAGAGTGTGAGGAGTGTGAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGAGTGTGCTGTGGAAGAGTGTGGAAGAGTGTGAGGAGTGTGAGGACTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAAGACCTGTAATGAAATTCCTGTGTGCAACCAAGTGTGTAAGGGAAAGAGACA 664
Db 887 ACCAGGAGGAAGACCTGTAATTCCTGTGTGCAACCAAGTGTGTAAGGGAAAGAGACA 946
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGCAAGAGGGGGAGATAGCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTGCGCAGAGCAGTGGTCCGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA 1006
Qy 725 AGAATCAAGAAAGAGCGCAGCGAGTCAAG-----GGCTCTCTCCAGTGGCAGCAGCA 778
Db 1007 AGAAGTCAAGAAAGAGCAGCAGCGGCGCAGGAGGCGGAGGAGTGTGAGCAGCAGCA 1066
Qy 779 AGCAGAGGAAGAACTGGGGGGCTTGGGGGAGGATGCCAAACCGCGAGAGGAGGAGGTG 838
Db 1067 AACAAAGGAAGAGCTGGGTGGCTTGGAGGAGACCCAGCGCCCGAGGAGATGAGGGA 1126
Qy 839 TGCAGAGGCAATCGCCCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCTTAGTG 898
Db 1127 TCCAGAGGCAATCCCTCTCCACACAGCCCGCTGATGAGCTGTGAGCCCGAGCAGCAT 1186
Qy 899 TCCTTGAATCAAGACCCCTGACTTTCAGAGCTTGGGA 934
Db 1187 CCTCTGCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

```

RESULT 9

```

US-10-176-918-135
; Sequence 135, Application US/10176918
; Publication No. US2003002725A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

```

; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 135
; LENGTH: 1675
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-918-135

```

```

Query Match 37.6%; Score 691.6; DB 5; Length 1675;
Best Local Similarity 84.5%; Pred. No. 7.1e-197;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

```

```

Qy 5 CGAGGAGAGGAGCGCCCGCGGTCCGCTCTGCTCTGGGTCCGGCTGGGGCCATGGAGTCCA 64
Db 287 CGCGGAGGAGGAGAACCGCCCGTCTTTAGGGTCCGGCCCGCGGCCCATGATTCAA 346
Qy 65 TGTCTGAGCTCGCGCCCGCGTCTCTTATTTCTTTGCTGTGCTGTGCTGTGCTGTGCTGTG 124
Db 347 TGCTGAGCGCGCTCCCGCTGTCTCTGCTTCTTCCCTTGTCTGTCTGTCTGTCTGTCTGC 406
Qy 125 TCCTTCTGCCCCGAGCTAGGCCCGAGTCCCGCGCGGCTGAGAGACCGACTGGGTGC 184
Db 407 TGCTGCGCGCCCGGAGCTGTGGCCCGAGCCAGCGCGGAGCTGAGAGAACACTGGGTTC 466
Qy 185 GATTGCCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTTG 244
Db 467 GCCTGCCCAGCAAAATGCGAAGTGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 526
Qy 245 AGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGAGCGGAGG 304
Db 527 AGGAAACGGGAAAGACCAAGGAAGTGTGACACCGGCTATGGCATCTCTGAGCAGAGG 586
Qy 305 GCTCTGAGTCAAGTACCAAGTCGAGCTTACCGTTAAATTTGAAGTCACTGAGACCATTT 364
Db 587 CCTCTGAGTCAAAATACCAAGTCGAGCTTTCGGTTAAATCGAAGTCACTGAGACCATTT 646
Qy 365 GCAAGAGGCTTCTGAGCTACAGCCTGCAACAGGAGAGGACTGGCAGCAACCGGTTTGCCA 424
Db 647 GCAAGAGGCTTCTGAGTATAGCTGTGCAACAGGAGAGGACCGCGCAATCGATTGCCA 706
Qy 425 AGGTATGTGCGAGACCTTTGAGAGCTGTGCAACACTTAGTGTGCAACAAAGGGGTCAAGTGG 484
Db 707 AGGGCATGTGAGAGACCTTTGAGACATTTACACAACTTGTGTATACAAAGGGGTCAAGTGG 766
Qy 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACTCTGAGAGAGTGGCTGACCTCAAGA 544
Db 767 TGATGGACATCCCTATGAGCTGTGGAACGAGACTCTTGCAGAGGTGGCTGACCTCAAGA 826
Qy 545 AGCAGTGTGAGCTGTGCTGTGGAAGAGTGTGGAAGAGTGTGAGGAGTGTGAGGACTGGTACAGGAACC 604
Db 827 AGCAGTGTGAGTGTGCTGTGGAAGAGTGTGGAAGAGTGTGAGGAGTGTGAGGACTGGTACAGGAACC 886
Qy 605 ACCAGGAGGAAGACCTGTAATTCCTGTGTGCAACCAAGTGTGTAAGGGAAAGAGACA 664
Db 887 ACCAGGAGGAAGACCTGTAATTCCTGTGTGCAACCAAGTGTGTAAGGGAAAGAGACA 946
Qy 665 CGAGTTGCTAGCAGAGCGGTGGTCTGCAAGAGGGGGAGATAGCCTCCCTGGGAGGGA 724
Db 947 CCAGTTGCTGCGCAGAGCAGTGGTCCGCAAGAGGGGAGACACAGCTGCCCTGGGAGGGA 1006

```

QY 725 AGAATCCAGGAAGCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGA 778
DB 1007 AGAAGTCAAGAAGAGAGCAGCGGGCCAGGCGAGCGAGGAGTAGCAGCAGA 1066
QY 779 AGCAGAGAAGAACTGGGGGGCTGGGGGAGGATGCCAAAGCGGAGGAGGAGGAGG 838
DB 1067 AACAAAGGAGAGCTGGGTGGCTTGAAGGAGACCCAGCCCCAGGAGGAGTGAAGGCA 1126
QY 839 TGCAGAAGGCATGCCCCCTCCACACAGCCCCCTGTGATGAGCTGTGAGCCAGCTTAGTG 898
DB 1127 TCCAGAAGGCATCCCCCTCTCACACACAGCCCCCTGTGATGAGCTGTGAGCCAGCAGCAT 1186
QY 899 TCCTTGAATCAGACCCCTGACTTCAGAGCTTGGGA 934
DB 1187 CCTCTGTCTGAGACCCCTGATTTTGAAGCTGAGGA 1222

RESULT 10

US-10-176-921-135

; Sequence 135, Application US/10176921
; Publication No. US2003002726A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C288

CURRENT APPLICATION NUMBER: US/10/176, 921

CURRENT FILING DATE: 2002-06-20

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 135

LENGTH: 1675

TYPE: DNA

ORGANISM: Homo Sapien

US-10-176-921-135

Query Match 37.6%; Score 691.6; DB 5; Length 1675;

Best Local Similarity 84.5%; Pred. No. 7.1e-197;

Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY 5 CGAGGAGAGAGCGCGCGGCTCGCTCTGCTGGGTCCGGCTGGGCCATGAGTGCA 64
DB 287 CGCGGAGAGAGAACCGCCCGCTCTTAGAGTCCGGCCCGCCGGCCCATGATTCAA 346
QY 65 TGCTGAGCTGGCGCCCGCTGCTCTTAATTTCTTTGCTGTGCTGTGCTGTGCTGTC 124
DB 347 TGCCTGAGCCCGGCTCGCTCTTCTGCTCTTCTTCTTCTTCTGCTGTGCTGTGCTGTC 406
QY 125 TCCTTCTTCCCGAGCTAGCGCCGAGTCCCGCGGGCTGAGAGACCGAGCTGGGTGC 184
DB 407 TGCTCCGGCCCGGAGCTGGGCCCGAGCGAGGAGCTGAGGAGAACGACTGGGTTC 466
QY 185 GATTGCCAGCAATGCGAGCTGCAAGTATGTTGCTGTGAGCTGAAGTCGGCTTTG 244
DB 467 GCCTGCCAGCAATGCGAGGTGTAATATGTTGCTGTGAGCTGAAGTCAGCCTTTG 526
QY 245 AGGAAACGGGAAAGAACCAAGGAAGTGATTGACACCGGCTATGGCATCTGGACGGGAAGG 304

DB 527 AGGAAACCGGCAAGCAAGGAGGTGATTGGCAGCGGCTATGGCATCTTGGACCAAGAGG 586
QY 305 GCTCTGGAGTCAAGTATACCAAGTCGGACTTACCGTTAATTTGAAGTCACTTGAGACCATTT 364
DB 587 CCTCTGGAGTCAAATATACCAAGTCGGACTTACCGTTAATTTGAAGTCACTTGAGACCATTT 646
QY 365 GCAAGAGCTTCTGGACTACAGCTGCAACAGGAGAGGAGTGGCAGCAACCGGTTTGCCA 424
DB 647 GCAAGAGCTTCTGGATTATAGCTTGCACAAAGGAGAGACCGGCAAGCAATCGATTTTGCCA 706
QY 425 AGGGTATCTCGAGAGCTTTTTCAGAGCTGTCACAACTAGTCCACAAAGGGGTCAAGGTGG 484
DB 707 AGGGCATGTTCAGAGACCTTTTTCAGACATTATACAACTGTGTACAAAGGGGTCAAGGTGG 766
QY 485 TGATGGATATCCCTTATGAGCTGTGGAAACGAGACCTTCAGCAGAGGTGTGCTCAAGCA 544
DB 767 TGATGGACATCCCTTATGAGCTGTGGAAACGAGACCTTCTGCAGAGGTGTGCTCAAGCA 826
QY 545 AGCAGTGTGAGCTGTGTTGGAAGTGTGGAAGGTTGGAAGGTTGATTCAGGAGTGTGTCAGCAACC 604
DB 827 AGCAGTGTGATGTGCTGTTGGAAGTGTGGAAGGTTGATTCAGGAGTGTGTCAGCAACC 886
QY 605 ACCAGGAGGAGAGCTGACTGACTGAATTCCTCTGTCGCCAACCAAGCTGTGAAGGGAAGGACA 664
DB 887 ACCAGGAGGAGAGCTGACTGACTGAATTCCTCTGTCGCCAACCAAGCTGTGAAGGGAAGGACA 946
QY 665 CGAGTTGCTTAGCAGAGCGGTGCTGTCGCAAGAGGGGACATAGCCTCCCTGGGAGGGA 724
DB 947 CCAGTTGCTGCGCAGAGCAGTGTGTCGCGCAAGAGGAGAGACACAGCTGCCCTGGGAGGA 1006
QY 725 AGAAATCCAAAGAAAGCGCAGCGAGTCAAG-----GGCTCTCCAGTGGCAGCAGA 778
DB 1007 AGAAGTCCAAAGAAAGAGAGCAGCGGGCCAGGCGAGGAGTAGCAGCAGA 1066
QY 779 AGCAGAGAAGAACTGGGGGGCTGGGGGAGGATGCCAAAGCGGAGGAGGAGGAGG 838
DB 1067 AACAAAGGAGAGCTGGGTGGCTTGAAGGAGACCCAGCCCCAGGAGGAGTGAAGGCA 1126
QY 839 TGCAGAAGGCATGCCCCCTCCACACAGCCCCCTGTGATGAGCTGTGAGCCAGCTTAGTG 898
DB 1127 TCCAGAAGGCATCCCCCTCTCACACACAGCCCCCTGTGATGAGCTGTGAGCCAGCAGCAT 1186

RESULT 11

US-10-137-865-135

; Sequence 135, Application US/10137865

; Publication No. US20030032155A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C154

; FILE REFERENCE: P3330R1C154

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2005, 23:12:51 ; Search time 931 Seconds
(without alignments)
1035.506 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3*
- 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691.6	37.6	1675	6	US-10-131-826A-135
2	104.4	5.7	946	6	US-10-750-185-37066
3	48.8	2.7	15376	7	US-11-121-086-5
4	42.6	2.3	168516	7	US-11-121-086-3
5	40.2	2.2	5468	6	US-10-821-234-49
6	40	2.2	1619	6	US-10-821-234-10
7	40	2.2	175416	7	US-11-121-086-43
8	39.2	2.1	191684	7	US-11-121-086-2
9	39	2.1	189539	7	US-11-121-086-16
10	37	2.0	1373	6	US-10-750-185-46147
11	37	2.0	1433	7	US-11-090-351-1
12	36.6	2.0	6691	7	US-11-186-284-3
13	36.6	2.0	1080000	6	US-10-928-446A-1
14	36.6	2.0	1080000	6	US-10-928-446A-181
15	36.6	2.0	1080000	6	US-10-928-446A-183
16	36.6	2.0	1080000	6	US-10-928-446A-185
17	36.6	2.0	1080000	6	US-10-928-446A-187
18	36.6	2.0	1080000	6	US-10-928-446A-189
19	36.6	2.0	1080000	6	US-10-928-446A-191
20	36.6	2.0	1080000	6	US-10-928-446A-193
21	36.6	2.0	1080000	6	US-10-928-446A-195
22	36.6	2.0	1080000	6	US-10-928-446A-197
23	36.6	2.0	1080000	6	US-10-928-446A-199

c	24	36.6	2.0	1080000	6	US-10-928-446A-201
	25	36.4	2.0	2023	6	US-10-995-561-55
	26	36.4	2.0	12591	6	US-10-995-561-13415
	27	36.4	2.0	14023	6	US-10-995-561-13221
	28	36.4	2.0	199130	6	US-10-995-561-13233
	29	36.2	2.0	1191	7	US-11-061-869-8
	30	36	2.0	1062	6	US-10-750-185-52971
	31	36	2.0	2132	6	US-10-750-185-46060
	32	36	2.0	3156	6	US-10-995-561-455
	33	36	2.0	3201	6	US-10-995-561-453
	34	35.8	1.9	1128	6	US-10-750-185-57964
	35	35.8	1.9	171936	6	US-10-933-025-24
	36	35.4	1.9	167116	7	US-11-121-086-44
	37	35.2	1.9	1758	6	US-10-821-234-637
	38	35.2	1.9	6786	7	US-11-069-834-59
	39	35	1.9	201	6	US-10-995-561-78059
	40	35	1.9	7666	7	US-11-069-834-51
	41	34.8	1.9	600	6	US-10-750-185-1127
	42	34.8	1.9	10129	7	US-11-044-111-21
	43	34.8	1.9	16964	6	US-10-995-561-13424
	44	34.8	1.9	51749	6	US-10-995-561-13245
	45	34.4	1.9	201	6	US-10-995-561-5006

ALIGNMENTS

RESULT 1

US-10-131-826A-135
; Sequence 135, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10131.826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588

Qy	899	TCCTTGAATCAAGACCCCTGACTTCTACAGAGCTTGGGA	933
Db	1187	CCTCTGTCTGTAGACCCCTGATTTTGAAGCTGAGGA	1222
RESULT 2			
US-10-750-185-37066			
; Sequence 37066, Application US/10750185			
; Publication No. US20050260603A1			
; GENERAL INFORMATION:			
; APPLICANT: MMI GENOMICS, INC.			
; APPLICANT: Denise, Sue K.			
; APPLICANT: KERR, Richard			
; APPLICANT: ROSENFELD, David			
; APPLICANT: HOLM, Tom			
; APPLICANT: BATES, Stephen			
; APPLICANT: FANTIN, Dennis			
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE			
; FILE REFERENCE: MM1100-2			
; CURRENT APPLICATION NUMBER: US/10/750,185			
; CURRENT FILING DATE: 2003-12-31			
; PRIOR APPLICATION NUMBER: US 60/437,482			
; PRIOR FILING DATE: 2002-12-31			
; NUMBER OF SEQ ID NOS: 64922			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 37066			
; LENGTH: 946			
; TYPE: DNA			
; ORGANISM: Bovine 19866880845079			
US-10-750-185-37066			
Query Match 5.7%; Score 104.4; DB 6;			
Best Local Similarity 84.8%; Pred. No. 1.9e-17;			
Matches 117; Conservative 0; Mismatches 21;			
Qy	199	TGCGAAGTGTCAAGTATGTTCTGTGTGGAGCTGAAGTCGG	CGG
Db	570	TCCCCAGTGTGTAATATGTTGCTGTGGAGCTGAAGTCAG	CAG
Qy	259	ACCAAGGAAGTGATTCAGCACCGGCTATGGCATCTCTGACG	CGG
Db	630	ACCAAGGAGTGTATCGACACACAGGCTATGGCATCTCTGAC	CGG
Qy	319	TACACCAAGTCGGACTTA	336
Db	690	TATACAAAGTCGTAAGTA	707
RESULT 3			
US-11-121-086-5			
; Sequence 5, Application US/11121086			
; Publication No. US20050266459A1			
; GENERAL INFORMATION:			
; APPLICANT: POULSEN, TIM S.			
; APPLICANT: NIELSEN, KIRSTEN V.			
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC			
; FILE REFERENCE: 09138.6000-00000			
; CURRENT APPLICATION NUMBER: US/11/121,086			
; CURRENT FILING DATE: 2005-05-04			
; PRIOR APPLICATION NUMBER: 60/567,570			
; PRIOR FILING DATE: 2004-05-04			
; NUMBER OF SEQ ID NOS: 107			
; SOFTWARE: PatentIn version 3.3			
; SEQ ID NO 5			
; LENGTH: 153376			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-11-121-086-5			
Query Match 2.7%; Score 48.8; DB 7; I			
Best Local Similarity 43.9%; Pred. No. 0.029;			
Matches 258; Conservative 0; Mismatches 327;			

	Query Match	2.3%	Score 42.6;	DB 7;	Length 168516;
	Best Local Similarity	47.5%;	Pred. No. 1.2;		
	Matches 160;	Conservative 0;	Mismatches 174;	Indels 3;	Gaps 1;
Qy	514	GAGACCTCAGAGGTGGCTGACCTCAGAACCACTGTGACGTGCTGCTGTGAAAGCTTT	573		
Db	68433	GTGCCTTAACTTGAGGAGGAGGAGGGGGAGGGGGAGAGAGGAGGAGGGGAGGAAGCG	68492		
Qy	574	GAGAGGTGATTGAGGACTGGTACAGAACCAACGAGGAGAGACCTGACTCAATTCCTC	633		
Db	68493	GCAGAGAGCTTTCAGGCGGAGGAGGAGGGGGAGGAGGGGGGAGGAGTAAGAGCTGGAA	68552		

	Query Match	2.2%;	Score 40.2;	DB 6;	Length 5468;
	Best Local Similarity	48.9%;	Pred. No. 1.4;		
	Matches 108;	Conservative	0;	Mismatches 113;	Indels 0;
	Gaps	0;			
Qy	643	CACGTGCTGAAGGGAAGACACGAGTTGCTCTTAGCAGACGGGTGCTGTGGCAAGAAGGGG	702		
Db	3054	CACGACGTGCAGGGCAAGATCCGGACTCTTCAGGAGCAGCTGGAGAATGGCCCCAACACG	3113		
Qy	703	GACATAGCTCCCTGGAGGGGAAGAAATCCAGAAAGACGACGCGGAGTCAAGGGCTCC	762		
Db	3114	CAGCTGGCCCGGCTGCAGCAGAGAACTCCATCTTCGGGATGCTTGAACGAGGCCACG	3173		
Qy	763	TCCAGTGTGGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCCCTGGGGGAGGATGCCAACGCC	822		
Db	3174	AGCCAGGTGGAGGACAAACAGAAACGAGAGCTGGCCAAAGCTTCGGCAGGAGCTCAGCAAG	3233		
Qy	823	GAGGAGGAGGGGTGTGCAGAAAGGCATGCCCTCCACACA	863		
Db	3234	GTCAAGAAAGAGCTGGTGAGAAAGTTCAGAGGCTGTGCGGCA	3274		

RESULT 6
US-10-821-234-10/c
; Sequence 10, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclerampsia
; FILE REFERENCE: 821A

```

; CURRENT APPLICATION NUMBER: US/10/821.234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 10
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-10

Query Match      2.2%; Score 40; DB 6; Length 1619;
Best Local Similarity 49.1%; Pred. No. 0.98;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 639 CAACACGTGTGAAGGAAAGGACACGAGTTGCTAGCAGAGCGGTGCTTGCAAGAA 698
Db 1418 CAGCAAGAAGAAGAGGGCTACAATGTGAACGACGAGAAAGCCAAAGGAGAAACAAGAA 1359

Qy 699 GGGGGACATACCTCCCTCTGGAGGGAGAAATCCAGNAGAGCGCAGCGGAGTCNAGGG 758
Db 1358 GGGCGAGGGCGCGACGGAAGAGGGAGGACCCCGAAGAGAGTGAAGCCCGAGCGGCC 1299

Qy 759 CTCCTCCAGTGCAGCAGCAAGCAGAGGAAGAACTGGGGGGCCTGGGGGAGGATGCCAA 818
Db 1298 CGCAGAGCCCGCGAGGCGCAAGAGGGGCAAGAGNAGGCCGACCAAGGACCGGAGGCAA 1239

Qy 819 CGCCGAGGAGGAGGAGGTGTGCAAGAGGCAATGCC 854
Db 1238 GGGCGAGGAGGAGGCGGAGGCGAGAAAGGACGCGCGCGC 1203

```

```

RESULT 7
US-11-121-086-43
; Sequence 43, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 43
; LENGTH: 175416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-43

Query Match      2.2%; Score 40; DB 7; Length 175416;
Best Local Similarity 47.0%; Pred. No. 5.7;
Matches 156; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

Qy 674 TAGCAGACGGTGTGCTGGCAAGAGGGGACATAGCCTCCCTGGAGGGGAAGAAATCCA 733
Db 60882 TGGAAATAGTGTGGGGAGGAATAAATCGGAAGAAGAACGCTGGGGAGAGATGGACGCTA 60941

Qy 734 AGAAGACCGCAGCGGAGTCAAGGCTCTCCAGTGGCA-CGACNAGCAGAGNAGAA 792
Db 60942 AGCAGAGAGATGGGGCGAGGAGAGGGGAACACAGCCACCGACAATGGAAGAGCAAGAA 61001

Qy 793 CTGGGGGCGCTGGGGGAGGATGCCAAACCGCAGGAGGAGGGGTGTGCAGAAAGCATCG 852
Db 61002 GAGCAGGAGAGCGTGAAGGGCTGAGAACCGAGCAGGCGAGGGGACTTCAGAGGCAGAA 61061

Qy 853 CCCCTCCACACAGCCCCCTGTAGTGTGAGCCCCAGCTTAGTGTCTTGAATCAAGA 912
Db 61062 GCGCTTCAGAGCTCTGCAAAATCCAAATGCTAAGAAAGAACATGCAGTCAATTTGACRAGG 61121

```

```

Qy 913 CCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCGCAGCTCCAGCAAGGACA 972
Db 61122 GCAGTTCAGAAATGCCACCGTGCACAGAGGATGGAAGCCAGGACCGCCGCCAAG 61181

Qy 973 GCTGCTCTCCAGCATCAGGCTCTCTCCCTTGG 1004
Db 61182 CAGGGCTGTGGGAGCTGGTGGCTCTCTCTGG 61213

RESULT 8
US-11-121-086-2
; Sequence 2, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 191684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-2

Query Match      2.1%; Score 39.2; DB 7; Length 191684;
Best Local Similarity 43.2%; Pred. No. 9.5;
Matches 185; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

Qy 425 AGGTATGTGCGAGACCTTTGAGACGCTTGACACACCTAGTCCACAAAGGGGTCAAGGTGG 484
Db 50266 AGGAGGTGGCGAGGTAGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50325

Qy 485 TGATGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
Db 50326 TGGAGGTGAAGGTGGAGGAGGTAGAGGTAGAGGTGGAGGTGGAGGTAAATGGAGGTGG 50385

Qy 545 AGCAGTGTGAGCTGCTGGTGAAGAGTTTGAAGAGGTCAITTAGGACTGTGTACAGGAACC 604
Db 50386 AAGTGGAGGTGGAGGTGATGTTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50445

Qy 605 ACCAGAGGAAAGACCTGACTGAATTCCTGTGTGCCAACACGCTCTGAAGGAAAGGACA 664
Db 50446 TGGAGGTGGTGGAGTTGCTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50505

Qy 665 CGAGTTCCCTAGCAGAGCGGTGCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGGA 724
Db 50506 AGGAGATGCTGGAGGTGGAGGTGATGAGATGGAAGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50565

Qy 725 AGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAGCAAGA 784
Db 50566 TGAAGGAAGAGGTCTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGG 50625

Qy 785 GGAAGGAAGTCTGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGAGGTGTGCAGA 844
Db 50626 TGAAGGAGGAGGTGGAGGTGCTGGAGGTGGTGGAGGTGGTGGAGGAGGAGGAGGTGTGGAGGTGGAGGTGG 50685

Qy 845 AGGCATCG 852
Db 50686 TGATATGG 50693

```

```

RESULT 9
US-11-121-086-16
; Sequence 16, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:

```



```

; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPW01-029F2RM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6691
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82)...(3717)
; US-11-186-284-3

Query Match          2.0%; Score 36.6; DB 7; Length 6691;
Best Local Similarity 58.9%; Pred. No. 13;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 52 GCCATGGAGTCATGCTCTGAGCTGCGCGCCGCTGCTTTATTTCTTTGCTGCTG 111
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79 GCCATGGATCCGCGCGGAGCGCTGCGCGCTGCTGCTGCGCGCTGCTGCTG 138

Qy 112 CTTCCGCTGCTCTTCTGTCGCGCGAGCTAGCGCGAGTCCGCG 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139 CTGTGCTGCTGCTGCGCGCGCTCTGTCGCGCGCGCGCGCGCG 185

RESULT 13
US-10-928-446A-1/c
; Sequence 1, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; TITLE OF INVENTION: VIRAL BUDDING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825234)
; OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825270)
; OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825401)
; OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825428)
; OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825473)
; OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825765)
; OTHER INFORMATION: a "c" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (825828)
; OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826041)
; OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826546)
; OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826654)
; OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826826)
; OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (826863)
; OTHER INFORMATION: the 'n' at position 826863 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (827008)
; OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (843055)
; OTHER INFORMATION: the 'n' at position 843055 may be 'g' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (843118)
; OTHER INFORMATION: the 'n' at position 843118 may be 't' or 'c'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871027)
; OTHER INFORMATION: the 'n' at position 871027 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871140)
; OTHER INFORMATION: the 'n' at position 871140 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (871168)
; OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872678)
; OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (872742)
; OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (925859)
; OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
```


OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (106117)..(106249)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1007860)..(1008036)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1010940)..(1011014)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018160)..(1018291)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1018800)..(1018919)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1020028)..(1020225)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1026659)..(1026736)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1028113)..(1028167)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1034316)..(1034374)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1041390)..(1041455)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1043121)..(1043350)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1044868)..(1044989)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1047519)..(1047589)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050296)..(1050391)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1060368)..(1060441)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062648)..(1062708)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1064561)..(1064620)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1066207)..(1066314)
OTHER INFORMATION: exon

FEATURE:
NAME/KEY: CDS
LOCATION: (1067768)..(1067864)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1068609)..(1068681)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1073289)..(1073388)
OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-181
Query Match 2.0%; Score 36.6; DB 6; Length 1080000;
Best Local Similarity 50.9%; Pred. No. 78;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Qy 1547 CTGGGTCTCATCAGCCTCCAGAGTGTCCCTCATCGATCTTTTGTGCTTTGTCCCA 1606
Db 236820 CTGGGTCTCATCAGCCTCCAGAGTGTCCCTCATCGATCTTTTGTGCTTTGTCCCA 236761
Qy 1607 ATCCAGGGCTGGAAGCCCATCACCATCATTTGGGGTAATCATCTACTGCTCTCAG 1717
Db 236760 AGTAGCTGGGATTACAGGCACCCACCACCATGCTGGCTAAGCAGTTCTTGGCTCAGCTCCA 236761
Qy 1667 GTGCTGGAGGCGCCGGGGTGTGGTGGGGTAATCATCTACTGCTCTCAG 1717
Db 236700 AGACTGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTTAG 236650
RESULT 15
US-10-928-446A-183/c
; Sequence 183, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; TITLE OF INVENTION: VIRAL BUDDING
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (826985)..(827017)
FEATURE:
OTHER INFORMATION: full exon 1 range is 826667-827008
FEATURE:
NAME/KEY: CDS
LOCATION: (843242)..(843315)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (922549)..(922630)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (926021)..(926059)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (929123)..(929176)
OTHER INFORMATION: exon

```

; NAME/KEY: CDS
; LOCATION: (993104)..(993154)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (999547)..(999608)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1000354)..(1000456)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1002118)..(1002284)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1006117)..(1006249)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1007860)..(1008036)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1010940)..(1011014)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018160)..(1018291)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018800)..(1018919)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020028)..(1020225)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1026659)..(1026736)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1028113)..(1028167)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1034316)..(1034374)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1041390)..(1041455)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1043121)..(1043350)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1044868)..(1044989)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1047519)..(1047589)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1050296)..(1050391)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (1060368)..(1060441)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1062648)..(1062708)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1064561)..(1064620)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1066207)..(1066314)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1067768)..(1067864)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1068609)..(1068681)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1073289)..(1073388)
; OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-183

Query Match      2.0%; Score 36.6; DB 6; Length 1080000;
Best Local Similarity 50.9%; Pred. No. 78;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 1547 CTGGGTCTCATCAGCCTCCAGAGTGTCCTCATCGATCTTTTGTGCTTGTGCTCCCA 1606
Db 236820 CTGGGTCTCATCAGCCTCCAGAGTGTCCTCATCGATCTTTTGTGCTTGTGCTCCCA 236761

Qy 1607 ATCCAGGGGCTGGAAGGCCATCACCATCATTTGAGGGTTAACTGTCAGTTACTAGGAG 1666
Db 236760 AGTAGCTGGGATTACAGGCACCCACCACCATGCTGGCTAAATTTTGTATTTTAGTAG 236701

Qy 1667 GTGCTGGGAGCGCCGGGTTGTTGGGTAATCACTCACTGGCTCTCAG 1717
Db 236700 AGACTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTTAG 236650

```

Search completed: December 30, 2005, 04:23:29
Job time : 941 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 03:38:26 ; Search time 9079 Seconds
(without alignments)
11513.945 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 5883141 seqs, 28421725653 residues

Word size : 0

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sy.*
- 12: gb_un.*
- 13: gb_vl.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1780	96.8	1882 9	BC013549 Mus muscu
2	1477	80.3	1909 9	AF361644 Mus muscu
3	577	31.4	1655 9	AK220209 Mus muscu
4	125	6.8	10892 14	AC117716 Mus muscu
5	99	5.4	89855 14	AC074220 Mus muscu
6	69	3.8	22256 14	AC134482 Rattus no
7	69	3.8	246404 14	AC131887 Rattus no
8	64	3.5	270958 14	AC135655 Rattus no
9	50	2.7	526 6	CQ921127 Sequence
10	50	2.7	580 6	AX198865 Sequence
11	50	2.7	580 6	AX209392 Sequence
12	50	2.7	612 8	AF161347 Homo sapi
13	50	2.7	904 8	BC022093 Homo sapi
14	50	2.7	1201 6	AR379595 Sequence
15	50	2.7	1277 8	BC008133 Homo sapi
16	50	2.7	1413 8	BC008898 Homo sapi
17	50	2.7	1413 8	BC008961 Homo sapi
18	50	2.7	1593 8	BC011767 Homo sapi

19	50	2.7	1675	6	AX464002	Sequence
20	50	2.7	1675	8	AY358960	Homo sapi
21	50	2.7	1720	8	BC004423	Homo sapi
22	50	2.7	4469	8	AK090425	Homo sapi
23	50	2.7	113109	8	HS475N16	Human DNA
24	49	2.7	2623	8	AK124196	Homo sapi
25	45	2.4	958	6	CQ716533	Sequence
26	45	2.4	962	8	HS080744	Homo sapien
27	36	2.0	1534	6	AR379729	Sequence
28	35	1.9	632	10	BV059085	SV212P6851
29	35	1.9	1657	6	BD194905	86 human
30	35	1.9	1657	6	CQ855233	Sequence
31	33	1.8	1658	6	BD218541	71 human
32	30	1.6	4881	9	AY267461	Mus muscu
33	30	1.6	165733	14	AC107706	Mus muscu
34	30	1.6	187371	34	AC164172	Mus muscu
35	30	1.6	212943	9	AC107787	Mus muscu
36	30	1.6	213550	9	AC153370	Mus muscu
37	29	1.6	103129	9	AL662809	Mus muscu
38	29	1.6	171820	9	AC126550	Mus muscu
39	29	1.6	200057	4	CR956379	Pig DNA s
40	29	1.6	216315	9	AL928963	Mouse DNA
41	29	1.6	258622	14	AC020877	Mouse DNA
42	28	1.5	1549	5	BC097110	Danio rer
43	28	1.5	114568	14	AC134480	Rattus no
44	28	1.5	148679	9	AC138308	Mus muscu
45	28	1.5	166302	5	CR385039	Zebrafish

ALIGNMENTS

BC013549 1882 bp mRNA linear ROD 28-SEP-2004
Mus musculus trinucleotide repeat containing 5, mRNA (cdna clone
MGC:19340 IMAGE:4222133), complete cds.

BC013549
BC013549
MGC.
MGC.

ORGANISM
Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 1882)

AUTHORS
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Sheezy, T.E., Brownstein, M.J., Ustin, T.B., Toshitoki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McKwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,
Schnerch, A., Schein, J.B., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1882)

Director MGC Project.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 25 Row: J Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 31982623.

FEATURES

source

1..1882
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:19340 IMAGE:4222133"
/tissue_type="Kidney, normal. 5 month old male mouse."
/clone_lib="NCI COAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

gene

1..1882
/gene="Thnc5"
/note="synonyms: CAG4A, ERDA5"
/db_xref="GeneID:72029"
/db_xref="MGI:1919279"
91..921
/gene="Thnc5"

CDS

/codon_start=1
/product="trinucleotide repeat containing 5"
/protein_id="AAH13549.1"
/db_xref="GI:15498836"
/db_xref="GeneID:72029"
/db_xref="MGI:1919279"
/translation="MESMSLAPRCLLPFLLLPLLLPAPKLGPSPAGAEEDWVR
LPSKEVCYVAVELKAPETGKTEVIDTGYGLDGKSGVKYFISDLRLIEVET
ICKLLDYSLKERGTGNRFAGKMSETFETLNLVHKGVKVMIDIPSLMNETSAEVA
DLKKQCDLVVEFEVIEDWYRNHQEEDLTFELCANHYLKGKDTSCLAERWSGKKGI
ASLGGKSKKRSYKVGSSSGSKQKELGLGEDANAEEBEGVQKASPLPHSPDPEL
"

ORIGIN

Query Match 96.8%; Score 1780; DB 9; Length 1882;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 GGAGGAAGCGCGCGGTCGCTGCTGCTGCGGTCGGGTCGGGCATGGAGTCATGTC 68
Db 45 GGAGGAAGCGCGCGGTCGCTGCTGCTGCGGTCGGGTCGGGCATGGAGTCATGTC 104

Qy 69 TGAGTCGCGCCCGCTCTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128
Db 105 TGAGTCGCGCCCGCTGCTATTTCTTTGCTGCTGCTGCTGCTGCTGCTGCT 164

Qy 129 TCCTGCCCGAGCTAGCGCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTCGATT 188
Db 165 TCCTGCCCGAGCTAGCGCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTCGATT 224

Qy 189 CCCAGCAAAATGCGAAGTGTGAAGTATGTTGCTGTGAGTGAAGTCCGGCTTTGAGGA 248
Db 225 CCCAGCAAAATGCGAAGTGTGAAGTATGTTGCTGTGAGTGAAGTCCGGCTTTGAGGA 284

Qy 249 AACGGGAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCTGGACGGAGGGCTC 308

Db 285 AACGGGAAGACCAAGGAAGTGATTGACACCGGCTATGGCATCTCTGGACGGAGGGCTC 344
Qy 309 TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATGAAGTCACTGAGACCATTTGCAA 368
Db 345 TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAAATGAAGTCACTGAGACCATTTGCAA 404
Qy 369 GAGGCTTCTGACTACAGCTTGACCAAGAGGAGGACTGGCAGCAACCGGTTTGCCAAAGG 428
Db 405 GAGGCTTCTGACTACAGCTTGACCAAGAGGAGGACTGGCAGCAACCGGTTTGCCAAAGG 464
Qy 429 TATGTCGAGACCTTTGAGACGCTGCACAACTTAGTCCAAAGAGGGTCAAGGTGGTAT 488
Db 465 TATGTCGAGACCTTTGAGACGCTGCACAACTTAGTCCAAAGAGGGTCAAGGTGGTAT 524

Qy 489 GGATATCCCTTATGAGCTGTGGAAGGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 548
Db 525 GGATATCCCTTATGAGCTGTGGAAGGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGCA 584
Qy 549 GTGTGACGCTGCTGCTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACACCA 608
Db 585 GTGTGACGCTGCTGCTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACACCA 644
Qy 609 GGAGGAAGACCTGACTGAATTCCTCTGTGTCACCAACGCTGTGAAGGAAAGGACACGAG 668
Db 645 GGAGGAAGACCTGACTGAATTCCTCTGTGTCACCAACGCTGTGAAGGAAAGGACACGAG 704
Qy 669 TTGCTTAGCAGAGCGGTGCTGCGCAAGAGGGGGACATAGCCTCCCTGGGAGGGAAGAA 728
Db 705 TTGCTTAGCAGAGCGGTGCTGCGCAAGAGGGGGACATAGCCTCCCTGGGAGGGAAGAA 764

Qy 729 ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAGCAGAGGAA 788
Db 765 ATCCAAGAAGACGCGAGCGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAGCAGAGGAA 824
Qy 789 GGAATCTGGGGGCTCTGGGGAGGATGCCAACCGCGAGAGAGGAGGTGTGCAGAAAGC 848
Db 825 GGAATCTGGGGGCTCTGGGGAGGATGCCAACCGCGAGAGGAGGAGGTGTGCAGAAAGC 884

Qy 849 ATCGCCCTCCACACAGCCCCCTTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 908
Db 885 ATCGCCCTCCACACAGCCCCCTTGATGAGCTGTGAGCCAGCTTAGTGTCTTGAATC 944
Qy 909 AAGACCCCTGACTTTCAGAGCTTTGGGACACGACAGCGCAGCGCAGCTCCAGCAAG 968
Db 945 AAGACCCCTGACTTTCAGAGCTTTGGGACACGACAGCGCAGCGCAGCTCCAGCAAG 1004
Qy 969 GACAGCTGCTGCAGCATCAGGTCCTCTCCCTTGGGTGTGCCCCCTTTCCTTCCCTTGA 1028
Db 1005 GACAGCTGCTGCAGCATCAGGTCCTCTCCCTTGGGTGTGCCCCCTTTCCTTCCCTTGA 1064

Qy 1029 CAACAGCAGAGGTGGAGGATCTGGGTGCTGGGAGACGGCACCCCAAGGGAAGAGGA 1088
Db 1065 CAACAGCAGAGGTGGAGGATCTGGGTGCTGGGAGACGGCACCCCAAGGGAAGAGGA 1124
Qy 1089 GGAGGAGCAGAAAGCAGCTCTCTTTCTACACAGTCCCCCTCAACGAGCTCCGGGGTCCACC 1148
Db 1125 GGAGGAGCAGAAAGCAGCTCTCTTTCTACACAGTCCCCCTCAACGAGCTCCGGGGTCCACC 1184
Qy 1149 CAGCATCCCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCAAT 1208
Db 1185 CAGCATCCCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGAGCATGAGGCAAT 1244
Qy 1209 AAGATGCTACACAGGCCCCCTTACGACGAGGAAGACTCCGTGAGCCTCAGCAGCCAGG 1268
Db 1245 AAGATGCTACACAGGCCCCCTTACGACGAGGAAGACTCCGTGAGCCTCAGCAGCCAGG 1304
Qy 1269 CCTGCTCTTCTTCCCAAGCAATTCCTTCTGCTGGTCTCTTGTGCGATGGTAAATTCG 1328
Db 1305 CCTGCTCTTCTTCCCAAGCAATTCCTTCTGCTGGTCTCTTGTGCGATGGTAAATTCG 1364
Qy 1329 AGAATCTTCAGGACAAACTCGGGTGTGCAAAAAGGGGCTGGACCCAGAGCCAGAGCCA 1388

Qy	909	AAGACCCCTGACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCCAGCAAG	968
Db	943	AAGACCCCTGACTTCAGAGCTTGGGACAGCACA-CGACAGCGCAGCGCAGCTCCAGCAAG	1001
Qy	969	GACAGCTGCTGTCCAGCATCAGGTCCTCCCTGGCTGTGCCCTTTCTTCCTCGTAA	1028
Db	1002	GACAGCTGCTGTCCAGCATCAGGTCCTCCCTGGCTGTGCCCTTTCTTCCTCGTAA	1061
Qy	1029	CAACAGCAAGGTGGAGGATCTGGGCTGTGGGAGCGCACCCCAAGGGAAGAGGA	1088
Db	1062	CAACAGCAAGGTGGAGGATCTGGGCTGTGGGAGCGCACCCCAAGGGAAGAGGA	1121
Qy	1089	GGAGGAGCAGAGGAGCTCTCTTCTACACAGTCCCTCCCTCAGCAGCTCCGCGGTCCACC	1148
Db	1122	GGAGGAGCAGAGGAGCTCTCTTCTACACAGTCCCTCCCTCAGCAGCTCCGCGGTCCACC	1181
Qy	1149	CAGCATCCCCAGGCTGAGATCAGGCTCCTGACATGGAAGCTGAAGACATGAGGCACAT	1208
Db	1182	CAGCATCCCCAGGCTGAGATCAGGCTCCTGACATGGAAGCTGAAGACATGAGGCACAT	1241
Qy	1209	AAGATGCTACAGCGCCCTTCAGCAGGAAGGACTCCGTGCGAGCTCAGCAGCCAGG	1268
Db	1242	AAGATGCTACAGCGCCCTTCAGCAGGAAGGACTCCGTGCGAGCTCAGCAGCCAGG	1301
Qy	1269	CCTGCTCTTCTCCACCAAGCATTCCTTCTGCTGCTGCTTCTGCGATGTTAAATTCG	1328
Db	1302	CCTGCTCTTCTCTCCACCAAGCATTCCTTCTGCTGCTGCTTCTGCGATGTTAAATTCG	1361
Qy	1329	AGAACTTCCAGGACAAAATCGGGGTGTGGCACAAAGGGGCTGGACCCAGAGCCAGAGCCA	1388
Db	1362	AGAACTTCCAGGACAAAATCGGGGTGTGGCACAAAGGGGCTGGACCCAGAGCCAGAGCCA	1421
Qy	1389	CGCCAGAGACTGCAGAGAGGGACCTGACCTAAACCCCTCGGAAAGCCAACTCTGCAGTTC	1448
Db	1422	CGCCAGAGACTGCAGAGAGGGACCTGACCTAAACCCCTCGGAAAGCCAACTCTGCAGTTC	1481
Qy	1449	CCGTGTCCACCCACTCCTCTGAGGAGCGCTCATGCTCTGCGCCAGCCCTTCTCCAGGGC	1508
Db	1482	CCGTGTCCACCCACTCCTCTGAGGAGCGCTCATGCTCTGCGCCAGCCCTTCTCCAGGGC	1541
Qy	1509	TACCAGAGTAAACACTTTTGGCTTTTCGGTTTGGTTCTCGTGGTCTCATCAGCCTCCAG	1568
Db	1542	TACCAGAGTAAACACTTTTGGCTTTTCGGTTTGGTTCTCGTGGTCTCATCAGCCTCCAG	1601
Qy	1569	AGTGTCCTCATCGATCTTTTGGCTTTTGGCTTTGTCCTTGTGCCCCCA	1606
Db	1602	AGTGTCCTCATCGATCTTTTGGCTTTTGGCTTTGTCCTTGTGCCCCCA	1639
RESULT 3			
AK220209			
LOCUS	Mus musculus mRNA for mFLJ00338 protein.		
DEFINITION	AK220209		
ACCESSION	AK220209.1		
VERSION	FLI_CDNA.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
AUTHORS	Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Kitamura,H., Nagase,T., Ohara.O. and Koga,H.		
TITLE	Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs Identified by Screening of Terminal sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries.		
JOURNAL	Published Only in Database (2005)		
REFERENCE	2		
AUTHORS	Okazaki,N., F.Kikuno,R., Nagase,T., Ohara,O. and Koga,H.		
TITLE	Direct Submission		

JOURNAL	Submitted (22-FEB-2005) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)		
FEATURES	Location/Qualifiers		
source	1..1655		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/clone="mek03217"		
	/tissue_type="adult spleen"		
gene	/note="vector:modified pbc sk+"<1..1655		
	/gene="mFLJ00338"		
CDS	<1..732		
	/gene="mFLJ00338"		
	/note="CDS is predicted by in silico analysis. Start codon is not identified."		
	/codon_start=1		
	/evidence=not experimental		
	/product="mFLJ00338 protein"		
	/protein_id="BAD90134.1"		
	/db_xref="GI:60359830"		
	/translation="AGAEETDWRLPSKCEVKYVAVELKSAFETGKTKEVIDTGYG ILDGSGVKYTKSLRLIEVTETICKRLDYSLHKERTGNSRFAGKMSFTFTLHNL VHKGKVVMDIPYELWNETSADVADKKQCDVLVEEFEEVIEDWYRNHQEDLTFELC ANHYLKGKDTSCLAERWSGKKGIASLGGKKKKRSVKSGSSSSSKQKRELGLGE DANAEEREGVQKLAERSLPHSPDEL"		
ORIGIN			
Query Match	31.4%;	Score 577;	DB 9; Length 1655;
Best Local Similarity	99.7%;	Pred. No. 0;	
Matches	677;	Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Qy	157	GCCGGGGCTGAGGAGACCGACTGGGTGCGATTGGCCAGCAAATGCGAAGTGTGCAAGTAT	216
Db	1	GCCGGGGCTGAGGAGACCGACTGGGTGCGATTGGCCAGCAAATGCGAAGTGTGCAAGTAT	60
Qy	217	GTTCTGTGGAGCTGAAGTTCGGCTTTTGGAGAAACAGGAAAGACCAGAAAGTGTATTCAC	276
Db	61	GTTCTGTGGAGCTGAAGTTCGGCTTTTGGAGAAACAGGAAAGACCAGAAAGTGTATTCAC	120
Qy	277	ACCGGCTATGCGATCCTGGACGGGAAGGGCTCTCGAGTCAAGTACACCAAGTCGGACTTGA	336
Db	121	ACCGGCTATGCGATCCTGGACGGGAAGGGCTCTCGAGTCAAGTACACCAAGTCGGACTTGA	180
Qy	337	CGGTTAATTTGAAGTCACTGAGACCAATTTGCAAGAGGCTTCTGGACTACAGCCTGCACAAG	396
Db	181	CGGTTAATTTGAAGTCACTGAGACCAATTTGCAAGAGGCTTCTGGACTACAGCCTGCACAAG	240
Qy	397	GAGAGGACTGGCAGCAACCGGTTTCCCAAGGGTATGTCGGAGACCTTTGAGACGCTGCAC	456
Db	241	GAGAGGACTGGCAGCAACCGGTTTCCCAAGGGTATGTCGGAGACCTTTGAGACGCTGCAC	300
Qy	457	AACCTAGTCCACAAAGGGGTCAAGGTGTGTGATGATATCCCTATGAGCTGTGGAACGAG	516
Db	301	AACCTAGTCCACAAAGGGGTCAAGGTGTGTGATGATATCCCTATGAGCTGTGGAACGAG	360
Qy	517	ACCTCAGCAGAGTGGCTGACCTCAAGAAGCAGTGTGACGTCGTGCTGGTGGAGAGAGTTGAA	576
Db	361	ACCTCAGCAGAGTGGCTGACCTCAAGAAGCAGTGTGACGTCGTGCTGGTGGAGAGAGTTGAA	420
Qy	577	GAGGTGATTGAGGACTGTGTACAGGAACCCACAGGAGGAAGACCTTGACTGAATTCCTCTGT	636
Db	421	GAGGTGATTGAGGACTGTGTACAGGAACCCACAGGAGGAAGACCTTGACTGAATTCCTCTGT	480
Qy	637	GCCAAACACGTGCTGAAGGGAAAGACACGAGTTGCTTCCTAGCAGAGCGGTGTCTGGCAAG	696
Db	481	GCCAAACACGTGCTGAAGGGAAAGACACGAGTTGCTTCCTAGCAGAGCGGTGTCTGGCAAG	540
Qy	697	AAGGGGACATAGAGCTTCCTCGGAGGGGAGAAATCCCAAGAAAGACCGCAGCGGAGTCAAG	756
Db	541	AAGGGGACATAGAGCTTCCTCGGAGGGGAGAAATCCCAAGAAAGACCGCAGCGGAGTCAAG	600

* 29635 30724: contig of 1090 bp in length
* 30725 30824: gap of 100 bp
* 30825 33991: contig of 1167 bp in length
* 31992 32091: gap of 100 bp
* 32092 33257: contig of 1166 bp in length
* 33258 33357: gap of 100 bp
* 33358 34530: contig of 1173 bp in length
* 34531 34630: gap of 100 bp
* 34631 35794: contig of 1164 bp in length
* 35795 35894: gap of 100 bp
* 35895 37060: contig of 1166 bp in length
* 37061 37160: gap of 100 bp
* 37161 38311: contig of 1151 bp in length
* 38312 38411: gap of 100 bp
* 38412 39569: contig of 1158 bp in length
* 39570 39669: gap of 100 bp
* 39670 40816: contig of 1147 bp in length
* 40817 40916: gap of 100 bp
* 40917 42058: contig of 1142 bp in length
* 42059 42158: gap of 100 bp
* 42159 43278: contig of 1120 bp in length
* 43279 43378: gap of 100 bp
* 43379 44477: contig of 1099 bp in length
* 44478 44577: gap of 100 bp
* 44578 45746: contig of 1169 bp in length
* 45747 45846: gap of 100 bp
* 45847 47007: contig of 1161 bp in length
* 47008 47107: gap of 100 bp
* 47108 48266: contig of 1159 bp in length
* 48267 48366: gap of 100 bp
* 48367 49521: contig of 1155 bp in length
* 49522 49621: gap of 100 bp
* 49622 50802: contig of 1181 bp in length
* 50803 50902: gap of 100 bp
* 50903 52050: contig of 1148 bp in length
* 52051 52150: gap of 100 bp
* 52151 53130: contig of 1163 bp in length
* 53131 53413: gap of 100 bp
* 53414 54553: contig of 1140 bp in length
* 54554 54653: gap of 100 bp
* 54654 55769: contig of 1116 bp in length
* 55770 55869: gap of 100 bp
* 55870 57009: contig of 1140 bp in length
* 57010 57109: gap of 100 bp
* 57110 58261: contig of 1152 bp in length
* 58262 58361: gap of 100 bp
* 58362 59517: contig of 1156 bp in length
* 59518 59617: gap of 100 bp
* 59618 60805: contig of 1188 bp in length
* 60806 60905: gap of 100 bp
* 60906 62065: contig of 1160 bp in length
* 62066 62165: gap of 100 bp
* 62166 63347: contig of 1182 bp in length
* 63348 63447: gap of 100 bp
* 63448 64612: contig of 1165 bp in length
* 64613 64712: gap of 100 bp
* 64713 65883: contig of 1171 bp in length
* 65884 65983: gap of 100 bp
* 65984 67149: contig of 1166 bp in length
* 67150 67249: gap of 100 bp
* 67250 68394: contig of 1145 bp in length
* 68395 68494: gap of 100 bp
* 68495 69619: contig of 1125 bp in length
* 69620 69719: gap of 100 bp
* 69720 70858: contig of 1139 bp in length
* 70859 70958: gap of 100 bp

Query Match 6.8%; Score 125; DB 14; Length 108992;
Best Local Similarity 100.0%; Pred. No. 4.7e-59;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 425 AGGGTATGTCGGAGACCTTTGAGACGCTTGACACCTAGTCACAAAGGGGTCAAGGTGG 484
|||||

Db 50305 AGGGTATGTCGGAGACCTTTGAGACGCTTGACACCTAGTCACAAAGGGGTCAAGGTGG 50364
QY 485 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 544
|||||
Db 50365 TGATGGATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGCTGACCTCAAGA 50424
QY 545 AGCAG 549
|||||
Db 50425 AGCAG 50429
|||||
RESULT 5
AC074220/c 89855 bp DNA linear HTG 15-MAY-2002
LOCUS Mus musculus strain C57BL6/J chromosome 17 clone RP23-76116,
DEFINITION WORKING DRAFT SEQUENCE, 38 unordered pieces.
AC074220
AC074220.7 GI:13876390
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 89855)
AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
REFERENCE 2 (bases 1 to 89855)
AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
TITLE Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
COMMENT On Apr 28, 2001 this sequence version replaced gi:11276104.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpgc@mendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 89115 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 13924: contig of 5942 bp in length
* 13925 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 19010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length

*	27078	27097: gap of unknown length
*	27098	28813: contig of 1722 bp in length
*	28820	28833: gap of unknown length
*	28840	31224: contig of 2385 bp in length
*	31225	31244: gap of unknown length
*	31245	35093: contig of 3849 bp in length
*	35094	35113: gap of unknown length
*	35114	38611: contig of 3498 bp in length
*	38612	38631: gap of unknown length
*	38632	42261: contig of 3630 bp in length
*	42262	42281: gap of unknown length
*	42282	44130: contig of 1849 bp in length
*	44131	44150: gap of unknown length
*	44151	47059: contig of 2909 bp in length
*	47060	47079: gap of unknown length
*	47080	49496: contig of 2417 bp in length
*	49497	49516: gap of unknown length
*	49517	52236: contig of 2720 bp in length
*	52237	52256: gap of unknown length
*	52257	54270: contig of 2014 bp in length
*	54271	54290: gap of unknown length
*	54291	56271: contig of 1981 bp in length
*	56272	56291: gap of unknown length
*	56292	58843: contig of 2558 bp in length
*	58850	58863: gap of unknown length
*	58870	60621: contig of 1752 bp in length
*	60622	60641: gap of unknown length
*	60642	62255: contig of 1614 bp in length
*	62256	62273: gap of unknown length
*	62276	63083: contig of 810 bp in length
*	63086	63105: gap of unknown length
*	63106	65440: contig of 2335 bp in length
*	65441	65460: gap of unknown length
*	65461	67623: contig of 2169 bp in length
*	67630	67643: gap of unknown length
*	67650	69178: contig of 1523 bp in length
*	69179	69198: gap of unknown length
*	69199	70095: contig of 897 bp in length
*	70096	70115: gap of unknown length
*	70116	71821: contig of 1706 bp in length
*	71822	71841: gap of unknown length
*	71842	73506: contig of 1665 bp in length
*	73507	73526: gap of unknown length
*	73527	74743: contig of 1223 bp in length
*	74750	74769: gap of unknown length
*	74770	75963: contig of 1194 bp in length
*	75964	75983: gap of unknown length
*	75984	77905: contig of 1922 bp in length
*	77906	77923: gap of unknown length
*	77926	79236: contig of 1311 bp in length
*	79237	79256: gap of unknown length
*	79257	81332: contig of 2076 bp in length
*	81333	81352: gap of unknown length
*	81353	83169: contig of 1817 bp in length
*	83170	83189: gap of unknown length
*	83190	84096: contig of 907 bp in length
*	84097	84116: gap of unknown length
*	84117	85743: contig of 1633 bp in length
*	85750	85769: gap of unknown length
*	85770	86770: contig of 1001 bp in length
*	86771	86790: gap of unknown length
*	86791	88207: contig of 1417 bp in length
*	88208	88227: gap of unknown length
*	88228	88976: contig of 749 bp in length
*	88977	88996: gap of unknown length
*	88977	88955: contig of 859 bp in length.

FEATURES

Query Match
5.4%; Score 99; DB 14; Length 89855;

```
Query Match      90.0%;   Score No. 22.7; Length Error;
Best Local Similarity 100.0%; Pred. No. 3.8e-44;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

9 GGAGGAAGCGCCGCCGGTCCGGCTCTGGTCTGGCTGGGCCATGGAGTCCATGTC 68

QY 676 GCAGAGCGTGTCTGGCAAGAGGGGACATAGCCTCCTGGGAGGAGGAATCCCAAG 735
 |||||
 Db 130160 GCAGAGCGTGTCTGGCAAGAGGGGACATAGCCTCCTGGGAGGAGGAATCCCAAG 130219
 |||||
 QY 736 AAGAAGCGC 744
 |||||
 Db 130220 AAGAAGCGC 130228

RESULT 7
 AC131887
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-3M14, WORKING DRAFT SEQUENCE, 3
 AC131887
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS_PASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus

REFERENCE
 AUTHORS
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
 1. (bases 1 to 246404)
 Muzny D.Marie, Metzker M.Lee, Abranzon S., Adams C., Alder J.,
 Allen C., Allen H., Albrooks S., Amin A., Anguiano D.,
 Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
 Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
 Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
 Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
 Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
 Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
 Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
 Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
 Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
 Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
 Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
 Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
 Grebergeorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W.,
 Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
 Hernandez R., Hines S., Hladun S.L., Hodgson A., Hogue M.,
 Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
 Karpach S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
 Kwis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
 Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
 Lorensuhewa L., Loulseghe H., Lozado R.J., Lu X., Ma J.,
 Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,
 Mangum B., Mapua P., Martin K., Martin R., Martinez E.,
 Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,
 Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
 Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,
 Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
 Nwaokelemeh O., Okwunonu G., Olarnpunsagoon A., Pal S., Parks K.,
 Pasternak S., Paul H., Perez A., Perez L., Pfamkoch C.,
 Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,
 Puazo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
 Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
 Rivers C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
 Sanders W., Savery G., Scherer S., Scott G., Shatman S., Shen H.,
 Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajda J.,
 Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
 Steimle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C.,
 Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
 Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,
 Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,
 Williams G., Willson R., Wleczyk R., Wooden H., Worley K.,
 Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
 Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
 Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
 Weinstock G. and Gibbs R.A.
 Direct Submission

TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
 2 (bases 1 to 246404)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (27-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 246404)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 9, 2003 this sequence version replaced gi:23270097.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBSQ

Center clone name: CH230-3M14

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 238268 bases at least Q40

Consensus quality: 240099 bases at least Q30

Consensus quality: 241516 bases at least Q20

Estimated insert size: 249377; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 161257: contig of 161257 bp in length

* 161258 161357: gap of unknown length

* 161358 188568: contig of 27211 bp in length

* 188569 188668: gap of unknown length

* 188669 246404: contig of 57736 bp in length.

Location/Qualifiers

1- 246404

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clones="CH230-3M14"

161258..161357

/estimated_length=unknown

161358..163910

/note="wgs contig"

164279..166935

/note="wgs contig"

188569..188668

/estimated_length=unknown

ORIGIN


```

Query Match          3.8%; Score 69; DB 14; Length 246404;
Best Local Similarity 100.0%; Pred. No. 6.2e-27;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 GCAGAGCGTGGTCTGGCAAGAGGGGACATACGCTCCCTGGAGGGAGAAATCCAG 735
    |||
Db 23303 GCAGAGCGTGGTCTGGCAAGAGGGGACATACGCTCCCTGGAGGGAGAAATCCAG 23362

QY 736 AAGAAGCGC 744
    |||
Db 23363 AAGAAGCGC 23371

RESULT 8
AC135655/c
LOCUS          AC135655          270958 bp      DNA      linear      HTG 20-NOV-2002
DEFINITION    Rattus norvegicus clone CH230-6908, *** SEQUENCING IN PROGRESS ***,
              4 unordered pieces.
ACCESSION     AC135655
VERSION       AC135655.2 GI:25138854
KEYWORDS      HTG; HTGS_PHASE3; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE        Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE     1 (bases 1 to 270958)
AUTHORS       Munzy,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
              Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
              Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
              Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
              Biewal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
              Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
              Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
              Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
              Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
              Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
              Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
              Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
              Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,I., Fan,G.,
              Fernandez,S., Finley,M., Flag,N., Forbes,L., Foster,M., Foster,P.,
              Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
              Gebregorgis,E., Gear,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
              Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
              Harvey,I., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,
              Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
              Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
              Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
              Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
              Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
              Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
              Lorusuwhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
              Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
              Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
              MaWhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
              Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
              Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,
              Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
              Nwankwaleme,O., Okwou,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
              Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
              Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
              Puzio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
              Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
              Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
              Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
              Sherty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
              Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
              Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
              Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,T., Usmani,K.,
              Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
              Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
              Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
              Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 270958)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270958)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 20, 2002 this sequence version replaced gi:24158508.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be whole
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCOMI
Center clone name: CH230-6908
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 206551 bases at least Q40
Consensus quality: 211776 bases at least Q30
Consensus quality: 215202 bases at least Q20
Estimated insert size: 208321; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 248554: contig of 248554 bp in length
* 248555 248654: gap of unknown length
* 248655 267829: contig of 19175 bp in length
* 267830 267929: gap of unknown length
* 267930 268946: contig of 1017 bp in length
* 268947 269046: gap of unknown length
* 269047 270958: contig of 1912 bp in length.

FEATURES Location/Qualifiers

source
1..270958
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-6908"
1..1141
/note="wgs_contig"
91405..94036
/note="wgs_contig"

misc_feature

misc_feature


```
misc_feature 102157..103191
/note="wgs_contig"
misc_feature 242793..243887
/note="wgs_contig"
misc_feature 246116..248554
/note="wgs_contig"
gap 248555..248654
/estimated_length=unknown
misc_feature 248655..249752
/note="wgs_contig"
gap 267830..267929
/estimated_length=unknown
gap 268947..269046
/estimated_length=unknown
ORIGIN
Query Match 3.5%; Score 64; DB 14; Length 270958;
Best Local Similarity 100.0%; Pred. No. 4.6e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 364 TCGAAGAGCTTCTGGACTACAGCTGCACAGGAGAGACTGGCAGCAACCGGTTTCC 423
Db 249480 TCGAAGAGCTTCTGGACTACAGCTGCACAGGAGAGACTGGCAGCAACCGGTTTCC 249421
QY 424 AAGG 427
Db 249420 AAGG 249417

RESULT 9
CQ921127 CQ921127 526 bp DNA linear PAT 23-NOV-2004
LOCUS Sequence 2327 from Patent WO2004097052.
DEFINITION CQ921127
ACCESSION CQ921127
VERSION CQ921127.1 GI:56211068
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Burczynski, M.B., Twine, N.C., Slonim, D.K., Trepicchio, W.L.,
Strahs, A., Immerman, F. and Dorner, A.J.
TITLE Methods for prognosis and treatment of solid tumors
JOURNAL Patent: WO 2004097052-A 2327 11-NOV-2004;
Wyeth (US); Burczynski, Michael E. (US)
FEATURES
source
1..526
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
misc_feature 265_
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 270
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 327
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 345
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 348..352
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 354..469
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 471
/note="n includes a, c, g, or t, or contains no
nucleotide"

misc_feature 102157..103191
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 242793..243887
/note="n includes a, c, g, or t, or contains no
nucleotide"
misc_feature 246116..248554
/note="n includes a, c, g, or t, or contains no
nucleotide"
gap 248555..248654
/estimated_length=unknown
misc_feature 248655..249752
/note="n includes a, c, g, or t, or contains no
nucleotide"
gap 267830..267929
/estimated_length=unknown
gap 268947..269046
/estimated_length=unknown
ORIGIN
Query Match 2.7%; Score 50; DB 6; Length 526;
Best Local Similarity 100.0%; Pred. No. 3.7e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 586 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 105 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 154

RESULT 10
AX198865 AX198865 580 bp DNA linear PAT 29-AUG-2001
LOCUS Sequence 1320 from Patent WO0151513.
DEFINITION AX198865
ACCESSION AX198865
VERSION AX198865.1 GI:15389191
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Algate, P.A.
TITLE Ovarian tumor-associated sequences
JOURNAL Patent: WO 0151513-A 1320 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..580
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 2.7%; Score 50; DB 6; Length 580;
Best Local Similarity 100.0%; Pred. No. 3.7e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 586 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 353 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 402

RESULT 11
AX209392
```


Homnidae; Homo.
1 (bases 1 to 1277)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shennen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R., Jordan, H., Moore, T., Max, S.I., Wang, J., Heise, F., Diatchenko, L., Marudina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavati, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S.C., Garcia, A.M., Gay, L.J., Hulyak, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, B., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smialowski, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 1277)
AUTHORS	Director MGC Project.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/

Contact: amg@bcm.tmc.edu
 Gundaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowals, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRAX Plate: 19 Row: n Column: 9.
 Location/Qualifiers
 1. .1277
 FEATURES
 source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="3457829"
/tissue_type="Cervix, carcinoma"
/clone_lib="NIH MGC_12"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

```

```

Query Match      2.7%; Score 50; DB 8; Length 1277;
Best Local Similarity 100.0%; Pred. No. 3.8e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 586 GAGGACTGTTACAGGAACACCAGGAGGAAGACCTGACTGAAATTCCTCTG 635
      |||||
Db 485 GAGGACTGTTACAGGAACACCAGGAGGAAGACCTGACTGAAATTCCTCTG 534

Search completed: December 30, 2005, 07:12:39
Job time : 9085 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 01:46:06 ; Search time 1051 Seconds
(without alignments)
11661.627 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacggaggaggagcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 4996997 seqs, 332346308 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9953994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_21:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004as:.*
13: Geneseqn2004bs:.*
14: Geneseqn2005s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1839	100.0	1839	4	AAS05278 Murine tr
2	1658	90.2	1848	4	AAS05286 DNA seque
3	1528	83.1	1553	14	ADV67441 Nucleotid
4	107	5.8	107	4	AAS05289
5	100	5.4	100	12	ADL88397 DNA up-re
6	100	5.4	100	12	ADL88396 DNA up-re
7	66	3.6	471	4	AAS05284 T243 gene
8	50	2.7	189	2	AAT23979 Human gen
9	50	2.7	526	13	ADU11888 Solid tum
10	50	2.7	580	4	AAS25051 Human ova
11	50	2.7	580	5	AAS25051 Human ova
12	50	2.7	901	8	ACC72109 Human NOV
13	50	2.7	934	12	ADM72164 Human NTR
14	50	2.7	1201	11	ADI30814 Human CDN
15	50	2.7	1201	13	ADS82881 Human lym
16	50	2.7	1226	13	ADU82624 Human MDD
17	50	2.7	1362	4	AAS05279 Human tri
18	50	2.7	1378	10	ACA54677 Human NF-
19	50	2.7	1378	14	ADU83175 Human tri

20	50	2.7	1596	8	ACC72108	Acc72108 Human NOV
21	50	2.7	1675	4	AAS21311	Aas21311 Human CDN
22	50	2.7	1675	8	ACA03670	Aca03670 cDNA enco
23	50	2.7	1675	8	ABX89208	Abx89208 DNA enco
24	50	2.7	1675	8	ACD41862	Acd41862 Human sec
25	50	2.7	1675	8	ACA04091	Aca04091 Human CDN
26	50	2.7	1675	9	ADA45654	Ada45654 Novel hum
27	50	2.7	1675	9	ADA76085	Ada76085 Human PRO
28	50	2.7	1675	9	ADA18735	Ada18735 Human PRO
29	50	2.7	1675	9	ADA61358	Ada61358 Homo sapi
30	50	2.7	1675	9	ADB19143	Adb19143 Novel hum
31	50	2.7	1675	9	ADB27684	Adb27684 cDNA enco
32	50	2.7	1675	9	ADA86163	Ada86163 Novel hum
33	50	2.7	1675	9	ADB15727	Adb15727 Human PRO
34	50	2.7	1675	9	ADA47513	Ada47513 Human PRO
35	50	2.7	1675	9	ADA67308	Ada67308 Human PRO
36	50	2.7	1675	9	ADB30315	Adb30315 cDNA enco
37	50	2.7	1675	9	ADA85611	Ada85611 Novel hum
38	50	2.7	1675	9	ADA96823	Ada96823 Human PRO
39	50	2.7	1675	9	ADA79127	Ada79127 Human PRO
40	50	2.7	1675	9	ADA87266	Ada87266 Novel hum
41	50	2.7	1675	9	ADB16468	Adb16468 Human PRO
42	50	2.7	1675	9	ADA91560	Ada91560 Novel hum
43	50	2.7	1675	9	ADB14623	Adb14623 Human PRO
44	50	2.7	1675	9	ADB18584	Adb18584 Novel hum
45	50	2.7	1675	9	ADA93799	Ada93799 Human PRO

ALIGNMENTS

RESULT 1
AAS05278
ID AAS05278 standard; cDNA; 1839 BP.

AC AAS05278;

XX 07-SEP-2001 (first entry)

DE Murine trinucleotide repeat protein (TRP) cDNA sequence.

KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; murine; ss.

OS Mus sp.

XX Key Location/Qualifiers

FT CDS 55..885

FT FT /*tag= a

FT FT /product= "TRP"

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US029382.

XX 26-OCT-1999; 99US-0161488P.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX P-PSDB; AAU02498.

XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
XX contains heterozygous disruption in a gene encoding TRP.

XX Claim 8; Fig 11; 106pp; English.

CC The present sequence encodes for murine trinucleotide repeat protein
 CC (TRP). The invention describes methods of producing embryonic stem (ES)
 CC cells comprising a heterozygous disruption in a target DNA sequence
 CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
 CC comprising a homozygous disruption in a gene encoding TRP, where the
 CC disruption inhibits the production of the wild type TRP. The invention
 CC also relates to identifying agents capable of affecting a phenotype of a
 CC knockout mouse. Also described are methods of determining whether
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
 CC phenotypic change. The transgenic animals and the cells are useful for
 CC identifying compounds capable of ameliorating disease symptoms, and as
 CC test substrates for the identification of drugs, pharmaceuticals,
 CC therapies and interventions which may be effective in treating
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
 CC disease. The animal models for trinucleotide repeat disorders are ideal
 CC model systems to study the progression of disease in vivo, the molecular
 CC basis of these diseases and show the features observed in human disease.
 CC Using the mice, it is possible to model both the pathogenic mechanism and
 CC the trinucleotide repeat instability in the mouse

XX
 SQ Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 U; 0 Other;

Query Match 100.0%; Score 1839; DB 4; Length 1839;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGAGGAGGAGCGCGCGGCTCGCTCTGCTCGGTCGGGTCCGGTGGCCATGGAG 60
 DB 1 GGCACGAGGAGGAGCGCGCGGCTCGCTCTGCTCGGTCGGGTCCGGTGGCCATGGAG 60
 QY 61 TCCATGCTGAGCTCGCGCGCGCTGCTCTTATTTCTTGTGCTGCTGCTCGCTG 120
 DB 61 TCCATGCTGAGCTCGCGCGCGCTGCTCTTATTTCTTGTGCTGCTGCTGCTG 120
 QY 121 CTGCTCTCTTCTCCCGAAGCTAGGCGCGGCTCCCGCGGCTGAGGAGACCGACTGG 180
 DB 121 CTGCTCTCTTCTCCCGAAGCTAGGCGCGGCTCCCGCGGCTGAGGAGACCGACTGG 180
 QY 181 GTGCGATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCCGCT 240
 DB 181 GTGCGATTGCCAGCAATGCGAAGTGTGCAAGTATGTTGCTGTGGAGCTGAAGTCCGCT 240
 QY 241 TTTGAGGAAACGGGAAAGACCAAGGAAGTATTGACACCGGTATGGCATCTTGACCGG 300
 DB 241 TTTGAGGAAACGGGAAAGACCAAGGAAGTATTGACACCGGTATGGCATCTTGACCGG 300
 QY 301 AAGGCTCTGGAGTCAAGTACACCAAGTTCGGAAGTTCAGGTTAATGAACTCACTGAGACC 360
 DB 301 AAGGCTCTGGAGTCAAGTACACCAAGTTCGGAAGTTCAGGTTAATGAACTCACTGAGACC 360
 QY 361 ATTTCCAGAGGCTTCTGGACTACAGCTGCAAGGAGAGACTGGCAGCAACCGGTTT 420
 DB 361 ATTTCCAGAGGCTTCTGGACTACAGCTGCAAGGAGAGACTGGCAGCAACCGGTTT 420
 QY 421 GCCAAGGATATCTCGGAGACTTTGAGACGCTGCAACACCTAGTCCACAAAGGGGTCAAG 480
 DB 421 GCCAAGGATATCTCGGAGACTTTGAGACGCTGCAACACCTAGTCCACAAAGGGGTCAAG 480
 QY 481 GTGGTGATGGATATCCCTATGAGCTGTGGAAACGAGACTCAGCAGAGAGTGGCTGACCTC 540
 DB 481 GTGGTGATGGATATCCCTATGAGCTGTGGAAACGAGACTCAGCAGAGAGTGGCTGACCTC 540
 QY 541 AAGAAGCATGTGAGCTGTGTAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG 600
 DB 541 AAGAAGCATGTGAGCTGTGTAAGATTTGAAGAGGTGATTGAGGACTGGTACAGG 600
 QY 601 AACCAACGAGGAGAGACTGACTCAATTTCTCTGTGCAACACCTAGTCCAGGGAAG 660
 DB 601 AACCAACGAGGAGAGACTGACTCAATTTCTCTGTGCAACACCTAGTCCAGGGAAG 660
 QY 661 GACACGAGTGTCTAGCAGAGCGGTGGTCTGGCAAGAGAGGGGACATAGCCTCTCCCTGGGA 720
 DB 661 GACACGAGTGTCTAGCAGAGCGGTGGTCTGGCAAGAGAGGGGACATAGCCTCTCCCTGGGA 720

QY 721 GGGAGAAATCCAAAGAAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 780
 DB 721 GGGAGAAATCCAAAGAAAGACGCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 780
 QY 781 CAGAGGAAGAAATGGGGGGCTGGGGAGATGCCAAACGCGAGGAGAGAGGGGTGTG 840
 DB 781 CAGAGGAAGAAATGGGGGGCTGGGGAGATGCCAAACGCGAGGAGAGAGGGGTGTG 840
 QY 841 CAGAGGATCGCCCTCCCAACAGCGCCCTGATGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 900
 DB 841 CAGAGGATCGCCCTCCCAACAGCGCCCTGATGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 900
 QY 901 CTTCAATCAAGACCCCTGACTTCAAGCTTGGGACAGCGCACAGCGCGCAGCGCAGCT 960
 DB 901 CTTCAATCAAGACCCCTGACTTCAAGCTTGGGACAGCGCACAGCGCGCAGCGCAGCT 960
 QY 961 CCAGCAAGGACAGCTGTGTCAGCATAGGTCTCTCCCTTGGCTGTGTCCTTTTCTT 1020
 DB 961 CCAGCAAGGACAGCTGTGTCAGCATAGGTCTCTCCCTTGGCTGTGTCCTTTTCTT 1020
 QY 1021 CCCTTGAACACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGCGGCAACCCCAAGG 1080
 DB 1021 CCCTTGAACACAGCAAGAGGTGGAAGATCTGGGGTGTGGGAGCGGCAACCCCAAGG 1080
 QY 1081 GAAGAGGAGGAGGAGCAGAGGCGAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGG 1140
 DB 1081 GAAGAGGAGGAGGAGCAGAGGCGAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGG 1140
 QY 1141 GGTCCACCCAGCTCCCGAGCTGAGATCCAGGCTCTCTGATGAGTGAAGCTGAAGAGCATG 1200
 DB 1141 GGTCCACCCAGCTCCCGAGCTGAGATCCAGGCTCTCTGATGAGTGAAGCTGAAGAGCATG 1200
 QY 1201 AGGCACATAGATCTTCCAGCGCCCTTCCAGCGAGGAGGACTCGTGCAGCTCAG 1260
 DB 1201 AGGCACATAGATCTTCCAGCGCCCTTCCAGCGAGGAGGACTCGTGCAGCTCAG 1260
 QY 1261 CAGCAGCGCTCTCTTCTTCCCAAGCATTTCTTCTGCTGCTGCTTGTGCGATGG 1320
 DB 1261 CAGCAGCGCTCTCTTCTTCCCAAGCATTTCTTCTGCTGCTGCTTGTGCGATGG 1320
 QY 1321 TAAATTCGAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGGTGAGCGCAGAGC 1380
 DB 1321 TAAATTCGAGAACTTCCAGGACAACTCGGGTGTGGCAAAAGGGGTGAGCGCAGAGC 1380
 QY 1381 CAGAGCCACGCGAGAGCTGCGAGAGGCACTGACCTAACCCCTTGGAAAGCCAATC 1440
 DB 1381 CAGAGCCACGCGAGAGCTGCGAGAGGCACTGACCTAACCCCTTGGAAAGCCAATC 1440
 QY 1441 TGCAAGTTCCTGTCACCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCGCTTCT 1500
 DB 1441 TGCAAGTTCCTGTCACCACTCTCTCTGAGGAGCGCTCATGCTCTGCCAGCGCTTCT 1500
 QY 1501 CCCAGGGCTACAGAGTAAACACCTTTTGGCTTTCGGTTTGGTTCCTGGGTCTCATCA 1560
 DB 1501 CCCAGGGCTACAGAGTAAACACCTTTTGGCTTTCGGTTTGGTTCCTGGGTCTCATCA 1560
 QY 1561 GCCTCCAGAGTGTCCCTCATCGATCTTTTGTGCTTGTCCCAATCCAGGGGTGG 1620
 DB 1561 GCCTCCAGAGTGTCCCTCATCGATCTTTTGTGCTTGTCCCAATCCAGGGGTGG 1620
 QY 1621 AAGGCATACCATCATTTGAGGCTTAACTGTGAGTCTAGGAGGTGCTGGGAGCGCC 1680
 DB 1621 AAGGCATACCATCATTTGAGGCTTAACTGTGAGTCTAGGAGGTGCTGGGAGCGCC 1680
 QY 1681 CGGGTGTGGTGGGTAATCATCTGCTCTCAGCTTCTAACACTGAGCGCCCTTA 1740
 DB 1681 CGGGTGTGGTGGGTAATCATCTGCTCTCAGCTTCTAACACTGAGCGCCCTTA 1740
 QY 1741 ATACAGTTCCTTCTGTTGTGCTGCTCCACGCCCCCACACACACCATATAATTTT 1800
 DB 1741 ATACAGTTCCTTCTGTTGTGCTGCTCCACGCCCCCACACACACCATATAATTTT 1800

QY 1801 CGATCTGTTTCATTAAGTGTAAAAA 1839
 DB 1801 CGATCTGTTTCATTAAGTGTAAAAA 1839

RESULT 2

AA05286
 ID AAS05286 standard; DNA; 1848 BP.

XX AC AAS05286;
 XX 07-SEP-2001 (first entry)

DE DNA sequence of an expanded T243 gene.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.

XX Mus sp.
 OS Synthetic.

XX Key Location/Qualifiers
 FT 55..894
 FT CDS
 FT /*tag= a
 FT /product= "Expanded TRP"

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US029382.

XX 26-OCT-1999; 99US-0161488P.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX P-PSDB; AAU02500.

XX Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP.

XX Disclosure; Fig 15; 106pp; English.

XX The present sequence represents the DNA sequence of an expanded T243
 CC gene. The T243 gene encodes for a trinucleotide repeat protein (TRP). The
 CC invention describes methods of producing embryonic stem (ES) cells
 CC comprising a heterozygous disruption in a target DNA sequence (preferably
 CC gene T243) encoding a TRP and of producing a knockout mouse comprising a
 CC homozygous disruption in a gene encoding TRP, where the disruption
 CC inhibits the production of the wild type TRP. The invention also relates
 CC to identifying agents capable of affecting a phenotype of a knockout
 CC mouse. Also described are methods of determining whether expansion of the
 CC trinucleotide repeat in a gene encoding TRP produces a phenotypic change.
 CC The transgenic animals and the cells are useful for identifying compounds
 CC capable of ameliorating disease symptoms, and as test substrates for the
 CC identification of drugs, pharmaceuticals, therapies and interventions
 CC which may be effective in treating trinucleotide repeat disorders e.g.
 CC fragile X syndrome and Huntington's disease. The animal models for
 CC trinucleotide repeat disorders are ideal model systems to study the
 CC progression of disease in vivo, the molecular basis of these diseases and
 CC show the features observed in human disease. Using the mice, it is
 CC possible to model both the pathogenic mechanism and the trinucleotide
 CC repeat instability in the mouse

XX Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 U; 0 Other;

Query Match 90.2%; Score 1658; DB 4; Length 1848;
 Best Local Similarity 100.0%; Pred. No. 0;

	Matches	1658;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	182	TGCGATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAAGTGGCTT	241							
DB	191	TGCGATTGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAAGTGGCTT	250							
QY	242	TTGAGGAAACGGGAAAGCAAGGAAAGTATTGACACCGGCTATGGCATCTCTGGAGGGA	301							
DB	251	TTGAGGAAACGGGAAAGCAAGGAAAGTATTGACACCGGCTATGGCATCTCTGGAGGGA	310							
QY	302	AGGGCTCTGGAGTCAAGTACACCAAGTCCGACTTACGGTTAATTAAGTCACTGAGACCA	361							
DB	311	AGGGCTCTGGAGTCAAGTACACCAAGTCCGACTTACGGTTAATTAAGTCACTGAGACCA	370							
QY	362	TTTGAAGAGGCTTCTGCACTACAGCTGCAACAGGAGGAGTGGCAGCAACCGGTTTG	421							
DB	371	TTTGAAGAGGCTTCTGCACTACAGCTGCAACAGGAGGAGTGGCAGCAACCGGTTTG	430							
QY	422	CCNAGGTTATGTCGGAGACCTTTGAGAGCTGCAACCTAGTCCACAAAGGGGTCNAGG	481							
DB	431	CCNAGGTTATGTCGGAGACCTTTGAGAGCTGCAACCTAGTCCACAAAGGGGTCNAGG	490							
QY	482	TGGTATGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCA	541							
DB	491	TGGTATGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACCTCA	550							
QY	542	AGAAAGCTGTGACGTGCTGGTGGAAAGTGTGAAAGAGTGTGAGGACTGGTACAGGA	601							
DB	551	AGAAAGCTGTGACGTGCTGGTGGAAAGTGTGAAAGAGTGTGAGGACTGGTACAGGA	610							
QY	602	ACCACAGGAGAGACCTGACTGAAATTCCTCTGTGTGCCAACACAGTGTGAAGGGAAGG	661							
DB	611	ACCACAGGAGAGACCTGACTGAAATTCCTCTGTGTGCCAACACAGTGTGAAGGGAAGG	670							
QY	662	ACAGAGTTGCTAGCAGAGCGGTGCTGTGCAAGAGGGGACATAGCTCCCTGGGAG	721							
DB	671	ACAGAGTTGCTAGCAGAGCGGTGCTGTGCAAGAGGGGACATAGCTCCCTGGGAG	730							
QY	722	GGAAAGAAATCCAAGAAAGCGCAGCGAGTCAAGGGCTCTCTCAGTGCGCAGCAGCAAGC	781							
DB	731	GGAAAGAAATCCAAGAAAGCGCAGCGAGTCAAGGGCTCTCTCAGTGCGCAGCAGCAAGC	790							
QY	782	AGAGAAAGAAATCCAAGAAAGCGCAGCGAGTCAAGGGCTCTCTCAGTGCGCAGCAGCAAGC	841							
DB	791	AGAGAAAGAAATCCAAGAAAGCGCAGCGAGTCAAGGGCTCTCTCAGTGCGCAGCAGCAAGC	850							
QY	842	AGAGGATCGCCCTCCACACAGCCCTGATGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCT	901							
DB	851	AGAGGATCGCCCTCCACACAGCCCTGATGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCT	910							
QY	902	TTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACACGACGAGCGCAGCGCAGCGAGCTC	961							
DB	911	TTGAATCAAGACCCCTGACTTTCAGAGCTTGGGACACGACGAGCGCAGCGCAGCGAGCTC	970							
QY	962	CAGCAAGACAGCTGTGCTCAGCATCAGGTCTCTCTCTCTGCTGTGCTGTGCTGTGCTGTGCT	1021							
DB	971	CAGCAAGACAGCTGTGCTCAGCATCAGGTCTCTCTCTCTGCTGTGCTGTGCTGTGCTGTGCT	1030							
QY	1022	CTTTGAAACAGCAAGAGGTGGAAGGATCTGGGGTGTGGGAGACGACCCCAAGGG	1081							
DB	1031	CTTTGAAACAGCAAGAGGTGGAAGGATCTGGGGTGTGGGAGACGACCCCAAGGG	1090							
QY	1082	AAGAGGAGGAGGAGCAGAGGAGCTCTCTTTTACACAGTCCCCCTCAGAGCTCCGGG	1141							
DB	1091	AAGAGGAGGAGGAGCAGAGGAGCTCTCTTTTACACAGTCCCCCTCAGAGCTCCGGG	1150							
QY	1142	GTCCACCCAGCATCCCCAGGCTGAGATCCAGGCTCTTCACTGGAAGCTGGAAGAGCATGA	1201							
DB	1151	GTCCACCCAGCATCCCCAGGCTGAGATCCAGGCTCTTCACTGGAAGCTGGAAGAGCATGA	1210							
QY	1202	GGCACAATGATGCTCACCAGGCCCCCTTACGCCAGGAGGAGCTCCCTGAGGCTCAGC	1261							
DB	1211	GGCACAATGATGCTCACCAGGCCCCCTTACGCCAGGAGGAGCTCCCTGAGGCTCAGC	1270							

QY 1262 AGCAGGCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCGATGGT 1321
 DB 1271 AGCAGGCTGCTCTTCTTCCACCAAGCATCTCTTCTGCTGCTCTTCTGCGATGGT 1330
 QY 1322 AAATTCGAGAACTTCCAGGACAAACTCGGGTGTGGCACAAGAGGGCTGGACCGCAGAGCC 1381
 DB 1331 AAATTCGAGAACTTCCAGGACAAACTCGGGTGTGGCACAAGAGGGCTGGACCGCAGAGCC 1390
 QY 1382 AGAGCCAGCCAGAGACTGCGAGAGGGCACTGACCTAAACCCCTTGGAAAGCAATCT 1441
 DB 1391 AGAGCCAGCCAGAGACTGCGAGAGGGCACTGACCTAAACCCCTTGGAAAGCAATCT 1450
 QY 1442 GCAGTTCCTGCTGCCACCACTCTCTCCAGAGAGCCCTCATGCTCTGCCAGCCCTTCTC 1501
 DB 1451 GCAGTTCCTGCTGCCACCACTCTCTCCAGAGAGCCCTCATGCTCTGCCAGCCCTTCTC 1510
 QY 1502 CCAGGGCTACAGAGTAACACCTTTTGGCCTTTGGTTCCTGGGTCTCATCAG 1561
 DB 1511 CCAGGGCTACAGAGTAACACCTTTTGGCCTTTGGTTCCTGGGTCTCATCAG 1570
 QY 1562 CTTGAGAGTGTCCCTCATGATCTTTTGGCTTTTGGCTTGTCCCAATCCAGGGGCTGGA 1621
 DB 1571 CTTGAGAGTGTCCCTCATGATCTTTTGGCTTTTGGCTTGTCCCAATCCAGGGGCTGGA 1630
 QY 1622 AGGCCATCACCATCATTTGGAGGCTTAACCTGTCTAGTTACTAGGAGGTCTGGAGCGGCC 1681
 DB 1631 AGGCCATCACCATCATTTGGAGGCTTAACCTGTCTAGTTACTAGGAGGTCTGGAGCGGCC 1690
 QY 1682 GGGGTGTTGGGTAACTCACTCACTGGCTCTCAGCCCTTCTAAGCTGACGCCCCCTTAA 1741
 DB 1691 GGGGTGTTGGGTAACTCACTCACTGGCTCTCAGCCCTTCTAAGCTGACGCCCCCTTAA 1750
 QY 1742 TACAGTTCCTTCTGTTGGTGAATCTCCAGCGCCCAACACACACCAATAAATTTATTC 1801
 DB 1751 TACAGTTCCTTCTGTTGGTGAATCTCCAGCGCCCAACACACACCAATAAATTTATTC 1810
 QY 1802 GATGCTGTTTCACTACTGTAACCAAAAAAAAAAAAAAAAAAAAA 1839
 DB 1811 GATGCTGTTTCACTACTGTAACCAAAAAAAAAAAAAAAAAAAAA 1848

RESULT 3

ADV67441
 ID ADV67441 standard; DNA; 1553 BP.
 XX AC ADV67441;
 XX DT 10-MAR-2005 (first entry)
 XX DE Nucleotide sequence of a protein involved in FGF signal transduction.
 XX KW fibroblast growth factor; FGF; signal transduction; organ formation;
 XX KW cerebellum formation; canceration mechanism; gene; ds.
 XX OS Mus musculus.
 XX Key Location/Qualifiers
 XX FT 72..902
 XX CDS /*tag= a
 XX /product= "protein involved in FGF signal transduction"
 XX WO2004108930-Al.
 XX 16-DEC-2004.
 XX 01-JUN-2004; 2004WO-JP007910.
 XX 05-JUN-2003; 2003JP-00161005.
 XX (RIKE) RIKEN KK.
 XX Okamoto H, Hirate Y;

XX WPI; 2005-048583/05.
 DR P-PSDB; ADV67442.
 XX Novel gene encoding protein involved in fibroblast growth factor signal transduction, useful for analyzing mechanism of formation of organs such as midbrain optic-tectum portion, cerebellum, inner ear and canceration mechanism of cell.
 XX Claim 2; SEQ ID NO 19; 91pp; Japanese.
 XX The present sequence encodes a protein involved in fibroblast growth factor (FGF) signal transduction. Such proteins are useful for analyzing the mechanism of formation of various organs, such as midbrain optic-tectum portion by a midbrain mesencephalon border, the mechanism of formation of cerebellum, mechanisms of formation of inner ear, blood vessels, limb buds, teeth, and canceration mechanisms of a cell.
 XX Sequence 1553 BP; 352 A; 445 C; 477 G; 279 T; 0 U; 0 Other;
 SQ
 Query Match 83.1%; Score 1528; DB 14; Length 1553;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 GGAGAAAGCGCCGCGGCTCGCTCTCTGCGTCCGGCTGGGCCATGGAGTCCATGTC 68
 DB 26 GGAGAAAGCGCCGCGGCTCGCTCTCTGCGTCCGGCTGGGCCATGGAGTCCATGTC 85
 QY 69 TGAGCTCGCGCCCGCTGCTCTTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 128
 DB 86 TGAGCTCGCGCCCGCTGCTCTTATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 145
 QY 129 TCCGTCGCCCAAGCTAGCCCGGCTGCTGCGGCTGAGGAGCCGACTGGGTGCGATT 188
 DB 146 TCCGTCGCCCAAGCTAGCCCGGCTGCTGCGGCTGAGGAGCCGACTGGGTGCGATT 205
 QY 189 GCCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTGGCTTTGAGGA 248
 DB 206 GCCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGAGCTGAAGTGGCTTTGAGGA 265
 QY 249 AACGGGAAAGACCAAGGAAGTGTGACACCGCTATGGCATCTGGACGGGAGGGCTC 308
 DB 266 AACGGGAAAGACCAAGGAAGTGTGACACCGCTATGGCATCTGGACGGGAGGGCTC 325
 QY 309 TGGAGTCAAGTACACCAAGTTCGAGCTTACGGTAAATTTGAAGTCACTGAGACATTTGCAA 368
 DB 326 TGGAGTCAAGTACACCAAGTTCGAGCTTACGGTAAATTTGAAGTCACTGAGACATTTGCAA 385
 QY 369 GAGGCTTCTGGACTACAGCCTGCAACAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGG 428
 DB 386 GAGGCTTCTGGACTACAGCCTGCAACAGGAGAGACTGGCAGCAACCGGTTTGCCAAAGG 445
 QY 429 TATGTGAGAGACCTTTGAGAGCTGCAACCTAGTCCAAAGGGGTCAAGGTGGTGTAT 488
 DB 446 TATGTGAGAGACCTTTGAGAGCTGCAACCTAGTCCAAAGGGGTCAAGGTGGTGTAT 505
 QY 489 GGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGTCACTCAAGAAACA 548
 DB 506 GGATATCCCTTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGTCACTCAAGAAACA 565
 QY 549 GTGTGAGCTGCTGCTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCA 608
 DB 566 GTGTGAGCTGCTGCTGGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCA 625
 QY 609 GGAGGAAAGACCTGATGAAATTTCTGCTGTGCCAAACCACTGTGAGGAAAGGACACAG 668
 DB 626 GGAGGAAAGACCTGATGAAATTTCTGCTGTGCCAAACCACTGTGAGGAAAGGACACAG 685
 QY 669 TTGCTAGCAGAGCGGTGCTGCTGCAAGAGGGGACATAGCTTCCCTGGGAGGGAAGAA 728
 DB 686 TTGCTAGCAGAGCGGTGCTGCTGCAAGAGGGGACATAGCTTCCCTGGGAGGGAAGAA 745
 QY 729 ATCCAAGAAAGCGCAGCGGAGTCAAGGGCTCTCTCCAGTGGCAGACAGCAGAGGAA 788


```
XX PN WO200303445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US014114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PA (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4790; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;
XX
XX Query Match 5.4%; Score 100; DB 12; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-25;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
XX Db 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60
XX
XX QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
XX Db 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100
XX
XX RESULT 6
XX ID ADL88396 standard; DNA; 100 BP.
XX AC ADL88396;
XX XX
XX DT 20-MAY-2004 (first entry)
XX DE DNA up-regulated in murine haematopoietic stem cells SeqID 4789.
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX KW HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; ds.
XX OS Mus sp.
XX XX
XX PN WO200303445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US014114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PA (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4790; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;
XX
XX Query Match 5.4%; Score 100; DB 12; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-25;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
XX Db 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60
XX
XX QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
XX Db 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100
XX
XX RESULT 6
XX ID ADL88396 standard; DNA; 100 BP.
XX AC ADL88396;
XX XX
XX DT 20-MAY-2004 (first entry)
XX DE DNA up-regulated in murine haematopoietic stem cells SeqID 4789.
XX KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
XX KW HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
XX KW common myeloid progenitor; CMP; bone marrow stem cell; mouse; ds.
XX OS Mus sp.
XX XX
XX PN WO200303445-A2.
XX PD 13-NOV-2003.
XX PF 05-MAY-2003; 2003WO-US014114.
XX PR 03-MAY-2002; 2002US-0377383P.
XX PA (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4790; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;
XX
XX Query Match 5.4%; Score 100; DB 12; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-25;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
XX Db 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60
XX
XX QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
XX Db 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100
XX
XX RESULT 7
XX ID AAS05284 standard; DNA; 471 BP.
XX AC AAS05284;
XX XX
XX DT 07-SEP-2001 (first entry)
XX DE T243 gene homologous sequence #1 generated by PCR.
XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX KW transgenic animal; knockout mouse; triplet repeat expansion;
XX KW fragile X syndrome; Huntington's disease; mouse; ds.
XX OS Mus sp.
XX XX
XX PN WO200130798-A1.
XX PD 03-MAY-2001.
XX PF 26-OCT-2000; 2000WO-US029382.
XX PR 26-OCT-1999; 99US-0161488P.
XX PA (DELT-) DELTAGEN INC.
XX PI Klein R, Matthews W, Moore M, Allen KD;
XX PR WPI; 2001-300473/31.
XX DR
```

```
PR 03-MAY-2002; 2002US-0377383P.
XX (STOW-) STOWERS INST MEDICAL RES.
XX PI Li L;
XX DR WPI; 2004-022656/02.
XX PT Classifying an unknown multi-lineage affiliated gene comprises isolating
XX PT expressed nucleic acid sequences from the discrete cell sub-populations.
XX PS Claim 7; SEQ ID NO 4789; 123pp; English.
XX CC This invention relates to a novel method for predicting gene potential by
XX CC associating nucleic acid sequences of unknown function with particular
XX CC sub-population profiles. Specifically, it refers to classifying an
XX CC unknown multi-lineage affiliated gene by collecting hybridisation data to
XX CC develop a gene expression map, in order to determine the discrete sub-
XX CC population where it is expressed. The present invention describes methods
XX CC for predicting the lineage commitment of genes associated with the self-
XX CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
XX CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors
XX CC (CLPs) and common myeloid progenitors (CMPs), which are collectively
XX CC referred to as bone marrow stem cells populations. As such, these methods
XX CC can be used to identify associated multi-lineage affiliated genes and
XX CC hence the underlying molecular mechanisms in physiological haematopoietic
XX CC development. This polynucleotide sequence is DNA associated with a murine
XX CC HSC sub population of cells of the invention.
XX SQ Sequence 100 BP; 28 A; 23 C; 31 G; 18 T; 0 U; 0 Other;
XX
XX Query Match 5.4%; Score 100; DB 12; Length 100;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-25;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 454 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 513
XX Db 1 CACACCTAGTCCACAAAGGGGTCAAGGTGGTGATATCCCTATGAGCTGTGGAAC 60
XX
XX QY 514 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 553
XX Db 61 GAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAGTGTG 100
XX
XX RESULT 7
XX ID AAS05284 standard; DNA; 471 BP.
XX AC AAS05284;
XX XX
XX DT 07-SEP-2001 (first entry)
XX DE T243 gene homologous sequence #1 generated by PCR.
XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX KW transgenic animal; knockout mouse; triplet repeat expansion;
XX KW fragile X syndrome; Huntington's disease; mouse; ds.
XX OS Mus sp.
XX XX
XX PN WO200130798-A1.
XX PD 03-MAY-2001.
XX PF 26-OCT-2000; 2000WO-US029382.
XX PR 26-OCT-1999; 99US-0161488P.
XX PA (DELT-) DELTAGEN INC.
XX PI Klein R, Matthews W, Moore M, Allen KD;
XX PR WPI; 2001-300473/31.
XX DR
```

XX Novel transgenic animals useful as animal model for characterization of
PT function of a gene encoding trinucleotide repeat proteins (TRPs),
PT contains heterozygous disruption in a gene encoding TRP.
XX Claim 11; Fig 14; 106pp; English.
XX
XX The present sequence for T243 homologous sequence #1 is generated by PCR
XX and is homologous to the T243 gene which encodes for a trinucleotide
XX repeat protein (TRP). The invention describes methods of producing
XX embryonic stem (ES) cells comprising a heterozygous disruption in a
XX target DNA sequence (preferably gene T243) encoding a TRP and of
XX producing a knockout mouse comprising a homozygous disruption in a gene
XX encoding TRP, where the disruption inhibits the production of the wild
XX type TRP. The invention also relates to identifying agents capable of
XX affecting a phenotype of a knockout mouse. Also described are methods of
XX determining whether expansion of the trinucleotide repeat in a gene
XX encoding TRP produces a phenotypic change. The transgenic animals and the
XX cells are useful for identifying compounds capable of ameliorating
XX disease symptoms, and as test substrates for the identification of drugs,
XX pharmaceuticals, therapies and interventions which may be effective in
XX treating trinucleotide repeat disorders e.g. fragile X syndrome and
XX Huntington's disease. The animal models for trinucleotide repeat
XX disorders are ideal model systems to study the progression of disease in
XX vivo, the molecular basis of these diseases and show the features
XX observed in human disease. Using the mice, it is possible to model both
XX the pathogenic mechanism and the trinucleotide repeat instability in the
XX mouse
XX Sequence 471 BP; 105 A; 129 C; 148 G; 88 T; 0 U; 1 Other;
XX
XX Query Match 3.6%; Score 66; DB 4; Length 471;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-13;
XX Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 9 GGAGGAGCGCGCGGCTCGCTCTGCTCTGGTCCGGCTGGGCGCATGGAGTCCATGTC 68
XX DB 406 GGAGGAGCGCGCGGCTCGCTCTGCTCTGGTCCGGCTGGGCGCATGGAGTCCATGTC 465
XX
XX QY 69 TGAGCT 74
XX DB 466 TGAGCT 471
XX
XX
XX RESULT 8
XX AAT23979
XX ID AAT23979 standard; cDNA to mRNA; 189 BP.
XX AC AAT23979;
XX DT 02-SEP-1996 (first entry)
XX DE Human gene signature HUMGS05940.
XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX KW human; cloning; mapping; non-biased library; diagnosis; detection;
XX KW cell typing; abnormal cell function; ss.
XX OS Homo sapiens.
XX PN W09514772-A1.
XX PD 01-JUN-1995.
XX PF 11-NOV-1994; 94WO-JP001916.
XX PR 12-NOV-1993; 93JP-00355504.
XX PA (MATS/) MATSUBARA K.
XX PA (OKUB/) OKUBO K.
XX PI Matsubara K, Okubo K;
XX

DR WPI; 1995-206931/27.
XX Single-stranded DNA for identifying gene signatures - isolated from 3'-
PT directed human cDNA library that reflects relative abundance of corresp.
PT mRNA in specific human tissues.
XX Claim 1; Page 1501; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp. double
XX -stranded DNA) which comprises one of the 7837 "GS" sequences given in
XX AA19001-T26937 and which is able to hybridise to part of human genomic
XX DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were
XX obtained from 3'-directed cDNA libraries prepared from various human
XX tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using
XX poly(T) as the sole primer. Since the 3'- untranslated sequence is unique
XX to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise
XX with specific mRNAs. Each library is constructed so as to reflect
XX accurately the relative abundance of different mRNAs in the particular
XX tissue from which it was derived. The appearance frequency of a given GS
XX in a cDNA library can be determined (esp. using primers and probes
XX derived from the GS sequences) as a means of diagnosing abnormal cell
XX function or for recognising different cell types
XX
XX Sequence 189 BP; 56 A; 44 C; 58 G; 21 T; 0 U; 10 Other;
XX
XX Query Match 2.7%; Score 50; DB 2; Length 189;
XX Best Local Similarity 100.0%; Pred. No. 6.2e-08;
XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 586 GAGGACTGGTACAGGAACCCAGGAGGAGACCTGACTGAATTCCTCTG 635
XX DB 5 GAGGACTGGTACAGGAACCCAGGAGGAGACCTGACTGAATTCCTCTG 54
XX
XX RESULT 9
XX ADU11888
XX ID ADU11888 standard; DNA; 526 BP.
XX AC ADU11888;
XX DT 27-JAN-2005 (first entry)
XX DE Solid tumour prognosis gene seqid 2327.
XX KW cytostatic; gene therapy; expression profile; solid tumour;
XX KW peripheral blood mononuclear cell; PBMC; prognosis; ds.
XX OS Unidentified.
XX PN W02004097052-A2.
XX PD 11-NOV-2004.
XX PF 29-APR-2004; 2004WO-US013587.
XX PR 29-APR-2003; 2003US-0466067P.
XX PR 23-JAN-2004; 2004US-0538246P.
XX PA (AMHP) WYETH.
XX PA (STRA) STRAHS A.
XX PI Strahs A, Trepicchio WL, Burczynski ME, Twine NC, Slonim DK;
XX PI Immerman F, Dorner AJ;
XX DR WPI; 2004-804779/79.
XX
XX A method, useful for prognosing and treating solid tumor, comprises
PT comparing an expression profile of a gene expressed in peripheral blood
PT mononuclear cells to a reference expression profile of a gene.
XX Disclosure; Page; 111pp; English.
XX The invention describes a method comprising comparing an expression

DR P-PSDB; ADM72134.

XX New human neurotransmission-associated proteins and polynucleotides for

PT diagnosing, preventing or treating diseases or conditions associated with

PT aberrant protein expression, e.g. cancer, hepatitis, AIDS, obesity or

PT stroke.

XX Claim 5; SEQ ID NO 45; 216pp; English.

PS The invention relates to human neurotransmission-associated proteins

XX (NTRAN) and encoding polynucleotides. The NTRAN polypeptides can be

CC expressed by standard recombinant methodology. The polypeptides and

CC polynucleotides are useful in diagnosing, preventing or treating diseases

CC or conditions associated with the decreased expression or overexpression

CC of NTRAN, such as autoimmune/inflammatory, cardiovascular, neurological,

CC developmental, cell proliferative, transport, psychiatric, metabolic or

CC endocrine disorders. These diseases may include AIDS, allergies, atopic

CC dermatitis, arthritis, thyroiditis, obesity, Parkinson's disease,

CC Alzheimer's disease, stroke, epilepsy, myocardial infarction,

CC hypertension, cancer, atherosclerosis or hepatitis. These are also useful

CC in assessing the effects of exogenous compounds on the expression of

CC nucleic acid and amino acid sequences of NTRAN. The NTRAN or its

CC fragments are useful in screening compounds for effectiveness as agonist

CC or antagonist of the polypeptides, or in altering the expression of the

CC target polynucleotide and compounds that specifically bind to or modulate

CC the activity of the polypeptide. The microarray is useful in monitoring

CC or measuring protein-protein interactions, drug-target interactions, and

CC gene expression profiles. The present sequence represents a specific

CC example of a polynucleotide encoding a human NTRAN polypeptide.

XX

SQ Sequence 934 BP; 232 A; 248 C; 283 G; 171 T; 0 U; 0 Other;

Query Match 2.7%; Score 50; DB 12; Length 934;

Best Local Similarity 100.0%; Pred. No. 4.7e-08;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 635

Db 595 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 644

RESULT 14

AD130814

ID AD130814 standard; cDNA; 1201 BP.

XX

AC AD130814;

XX

DT 17-JUN-2004 (first entry)

XX

DE Human cDNA #140.

XX

KW Human; gene; ss; immunological response; immunopathological condition;

KW Crohn's disease; asthma; ulcerative colitis; hyperesinophilia;

KW irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;

KW acute monocytic leukaemia; antiinflammatory; antiaesthetic; antiulcer;

KW osteopathic; antiarthritic; antirheumatic; cytostatic.

XX

OS Homo sapiens.

XX

PN US6607879-B1.

XX

PD 19-AUG-2003.

XX

PF 09-FEB-1998; 98US-00023655.

XX

PR 09-FEB-1998; 98US-00023655.

XX

PA (INCY-) INCYTE CORP.

XX

PI Cocks BG, Stuart SG, Seilhamer JJ;

XX

DR WPI; 2003-895307/82.

XX

PT A composition comprising a plurality of cDNAs, useful for detecting

PT altered expression of genes in an immunological response or for

PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma

PT or osteoarthritis.

XX Claim 1; SEQ ID NO 140; 50pp; English.

PS The invention relates to a composition comprising a plurality of cDNAs

XX for detecting the altered expression of genes in an immunological

CC response. The invention also relates to a method of diagnosing or

CC monitoring the treatment of an immunopathological condition in a sample,

CC comprising obtaining nucleic acids from a sample, contacting the nucleic

CC acids of the sample with an array comprising the plurality of cDNAs under

CC conditions to form one or more hybridisation complexes, detecting the

CC hybridisation complexes and comparing the levels of the detected

CC hybridisation complexes with the level of hybridisation complexes

CC detected in a non-diseased sample, where an altered level of the detected

CC hybridisation complexes correlates with the presence of an

CC immunopathological condition. Also disclosed are an expression profile

CC comprising a microarray and a plurality of detectable complexes and a

CC method for identifying a plurality of polynucleotide probes. The cDNAs

CC are useful as hybridisable array elements in a microarray for monitoring

CC the expression of target polynucleotides. The microarray can be used in

CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,

CC ulcerative colitis, hyperesinophilia, irritable bowel syndrome,

CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in

CC identifying agents for the treatment of the diseases. The microarray may

CC also be used in drug discovery and development, toxicological and

CC carcinogenicity studies, forensics or pharmacogenomics. The composition

CC may also be used in purification of a subpopulation of mRNAs, cDNAs or

CC genomic fragments. This sequence represents a human cDNA of the

CC invention. Note: The sequence data for this patent did not form part of

CC the printed specification but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html.

XX

SQ Sequence 1201 BP; 293 A; 342 C; 347 G; 219 T; 0 U; 0 Other;

Query Match 2.7%; Score 50; DB 11; Length 1201;

Best Local Similarity 100.0%; Pred. No. 4.5e-08;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 635

Db 421 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAAATTCCTCTG 470

RESULT 15

ADS82881

ID ADS82881 standard; cDNA; 1201 BP.

XX

AC ADS82881;

XX

DT 11-AUG-2005 (first entry)

XX

DE Human lymph node cDNA #140.

XX

KW ss; gene; human; immunological response; blood cell; cancer;

KW immunopathological; AIDS; allergy; anaemia; asthma; atherosclerosis;

KW bronchitis; ulcerative colitis; diabetes; multiple sclerosis;

KW osteoporosis; pancreatitis; infection; arthritis; lymph node.

XX

OS Homo sapiens.

XX

PN US2004077003-A1.

XX

PD 22-APR-2004.

XX

PF 14-AUG-2003; 2003US-00641643.

XX

PR 09-FEB-1998; 98US-00023655.

XX

PA (INCY-) INCYTE CORP.

XX

Search completed: December 30, 2005, 04:41:05
Job time : 1054 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 03:44:32 ; Search time 6729 Seconds
(without alignments)
12786.661 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggacaggaggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 41078325 seqs, 23393541228 residues

Word size : 0
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gsl1:*
10: gb_gsl2:*
11: gb_gsl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1785	97.1	1787	AK086940	Mus muscu
2	1690	91.9	1832	AK049358	Mus muscu
3	1683	91.5	1684	AK031742	Mus muscu
4	1609	87.5	1896	AK085617	Mus muscu
5	1528	83.1	1553	AK005532	Mus muscu
6	1277	69.4	1621	AK082340	Mus muscu
7	966	52.5	1376	AK082749	Mus muscu
8	871	47.4	1586	AK010702	Mus muscu
9	766	41.7	766	AK166520	HES2 41
10	760	41.3	766	BG974240	602843972
11	679	36.9	903	CB320984	AGENCOURT
12	663	36.1	926	BQ922065	AGENCOURT
13	662	36.0	892	CF553710	AGENCOURT
14	656	35.7	937	CB203728	AGENCOURT
15	637	34.6	699	CB208584	AGENCOURT
16	629	34.2	721	AY403772	MNS15855
17	626	34.0	627	BB531047	Mus muscu
18	608	33.1	643	BB661603	BB531047
19	601	32.7	625	BB633565	BB661603
20	588	32.0	641	BB637089	BB633565
21	586	31.9	586	BM232538	BB637089
22	582	31.6	656	BI082275	K0324G03- BI082275 602877456

23	572	31.1	737	2	BG873094	BG873094 602794679
24	570	31.0	696	6	CD353185	UI-M-GLO-
25	565	30.7	665	6	CF732471	UI-M-HAO-
26	558	30.3	656	2	BB504341	BB504341
27	556	30.2	644	1	BB190512	BB190512
28	549	29.9	768	5	BUS22275	AGENCOURT
29	547	29.7	639	8	CX217093	CX217093 MNS31093
30	542	29.5	758	7	CO801920	AGENCOURT
31	541	29.4	623	2	BB661026	BB661026 BB661026
32	539	29.3	836	2	BI078951	BI078951 602873203
33	539	29.3	1078	5	BQ714545	AGENCOURT
34	533	29.0	608	2	BB605812	BB605812
35	526	28.6	645	2	BB498921	BB498921
36	525	28.5	985	2	BF785328	BF785328 60211557
37	523	28.4	651	2	BB385488	BB385488 BB385488
38	522	28.4	573	1	AA260027	AA260027 va52c07.r
39	521	28.3	632	2	BB558801	BB558801
40	516	28.1	516	3	BM223931	K0143E06-
41	512	27.8	564	7	CN700300	E0433C02-
42	503	27.4	503	5	BU696150	BU696150 LL21n1070
43	497	27.0	497	2	BG062828	BG062828 L0958D12-
44	489	26.6	631	5	BY705743	BY705743
45	481	26.2	557	3	BM224022	BM224022 K0144G10-

ALIGNMENTS

RESULT 1	AK086940	1787 bp	mRNA	linear	HTC 03-APR-2004
LOCUS	Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched library, clone:R030013D02 product:1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.				
ACCESSION	AK086940				
VERSION	AK086940.1	GI:26352264			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	10349636				
REFERENCE	1				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
PUBMED	11042159				
REFERENCE	2				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Maemoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, W., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
PUBMED	11076861				
REFERENCE	3				
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409, 685-690 (2001)				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409, 685-690 (2001)				
PUBMED	11076861				
REFERENCE	5				

2
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE
AUTHORS 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
PUBMED 11076861
REFERENCE
AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 695-690 (2001)
REFERENCE
AUTHORS 5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS 6 (bases 1 to 1684)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
1. 1684
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6230400B22"
/db_xref="taxon:10090"
/clone="6230400B22"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="11 days embryo"
CDS
1. 757
/note="unnamed protein product; 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA

160025D17 GENE) (SPTR|Q9DAU1, evidence: FASTY, 100%ID, 100%length, match=828)
putative"
/protein id="BAC27534.1"
/db_xref="GI:26327581"
/translation="PAPKLGSPAGAEETDMVRLPSKCEVKYKVAVELKSAFEETGKT
KEVIDTGVILDKGSGVKYKSDRLIEVTETICKRLDLSLHKERTGNSRPAKMS
ETFEFLNLVHKGVKVMNDIPYELWNESAEDLVKQCDLVEFEVIEDWYRNHQ
BEDTEFLCANHLVKGDTSCLAERWSKGDIASLGKYSKKRSGVKSGSSGSKQ
RKELGLGEDANABEEGVQKASPLPHSPDEL"
ORIGIN
Query Match 91.5%; Score 1683; DB 4; Length 1684;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 CCTGCCCGAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGCGATTG 189
DB 2 CCTGCCCGAGTAGGCCCGAGTCCCGCGGGCTGAGGAGACCGACTGGGTGCGATTG 61
QY 130 CCCAGCAATGCAAGTAGTGCAAGTATGTGTGTGGAGCTGAAGTCCGGCTTTTGAAGAA 249
DB 62 CCCAGCAATGCAAGTAGTGCAAGTATGTGTGTGGAGCTGAAGTCCGGCTTTTGAAGAA 121
QY 250 ACGGAAAGACCAAGAAAGTAGTGACACCGGCTATGTGATCTCGACGGGAGGCTCT 309
DB 122 ACGGAAAGACCAAGAAAGTAGTGACACCGGCTATGTGATCTCGACGGGAGGCTCT 181
QY 310 GGAGTCAAGTACCAAGTCCGAGCTTACGGTTTATTTGAAGTCACTGAGACCAATTTGCAAG 369
DB 182 GGAGTCAAGTACCAAGTCCGAGCTTACGGTTTATTTGAAGTCACTGAGACCAATTTGCAAG 241
QY 370 AGGTTCTGGACTACAGCTCGACAAAGAGAGGAGCTGGCAGCAACCGGTTTGCACAGGCT 429
DB 242 AGGTTCTGGACTACAGCTCGACAAAGAGAGGAGCTGGCAGCAACCGGTTTGCACAGGCT 301
QY 430 ATGTGGAGACCTTTGAGACGCTGCACAACTAGTTCACAAAGGGGTCAAGGTGGTGATG 489
DB 302 ATGTGGAGACCTTTGAGACGCTGCACAACTAGTTCACAAAGGGGTCAAGGTGGTGATG 361
QY 490 GATATCCCTATGAGCTGTGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAG 549
DB 362 GATATCCCTATGAGCTGTGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAGCAG 421
QY 550 TGTGACGTGCTGTGGAGAGTGTGAGAGGTGATTGAGGACTGCTACAGAACCCACGAG 609
DB 422 TGTGACGTGCTGTGGAGAGTGTGAGAGGTGATTGAGGACTGCTACAGAACCCACGAG 481
QY 610 GAGGAAGACCTTGACTGAATTCCTCTGTGCCCAACCACTGCTGAAGGGGAAAGCAGAGT 669
DB 482 GAGGAAGACCTTGACTGAATTCCTCTGTGCCCAACCACTGCTGAAGGGGAAAGCAGAGT 541
QY 670 TGCCTAGCAGAGCGGTGTGTGGCAAGAGGGGAGACATAGGCTTCCTGGAGGGAGAAA 729
DB 542 TGCCTAGCAGAGCGGTGTGTGGCAAGAGGGGAGACATAGGCTTCCTGGAGGGAGAAA 601
QY 730 TCCAGAGAGAGCCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAG 789
DB 602 TCCAGAGAGAGCCGAGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAGCAGAGGAG 661
QY 790 GAACTGGGGGGCTTGGGGGAGGATGCCAAACCGCAGGAGGAGGGGTGTGCAGAGGCA 849
DB 662 GAACTGGGGGGCTTGGGGGAGGATGCCAAACCGCAGGAGGAGGGGTGTGCAGAGGCA 721
QY 850 TCGCCCTTCCACACAGCCCCCTGTAGTGTGTGAGCCAGCTTAGTGTCTTGAATCA 909
DB 722 TCGCCCTTCCACACAGCCCCCTGTAGTGTGTGAGCCAGCTTAGTGTCTTGAATCA 781
QY 910 AGACCCCTGACTTCAAGAGCTTGGGACAGCAGCGGAGCGGAGCTCCAGGAGTCCAGCAGG 969
DB 782 AGACCCCTGACTTCAAGAGCTTGGGACAGCAGCGGAGCGGAGCTCCAGGAGTCCAGCAGG 841

Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
<http://fantom.gsc.riken.jp/>
Location/Qualifiers

FEATURES

source
1..1621
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:C230039G09"
/db_xref="taxon:10090"
/clone="C230039G09"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
230..685
/note="unnamed protein product; 1600025D17RIK PROTEIN
(PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA
1600025D17 GENE) (SPTR|O9DAU1, evidence: FASTI, 100%ID,
100%length, match=828)
putative"
/codon_start=1
/protein_id="BAC38471.1"
/db_xref="GI:26349663"
/translation="MSETFETFLNLVHKVGVVMDIPYELWNETSAEVADLKKQCDVL
VEEFEEVIEDWYRKHQEDLTFELCANHLKQDTSCLAEWSKGGKDIASLGGKSK
KKRGVSGSSSSSKRKELGGLGEDANAEIEGVQKASPLHSPDDEL"

CDS

ORIGIN

Query Match 69.4%; Score 1277; DB 4; Length 1621;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1397; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	424	AAGGTATGTCGGAGACCTTTGAGACGCTGCACAACTAGTCCAAAGGGGTCAAGGTG	483
DB	224	AAGGTATGTCGGAGACCTTTGAGACGCTGCACAACTAGTCCAAAGGGGTCAAGGTG	283
QY	484	GTGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAG	543
DB	284	GTGATGGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAG	343
QY	544	AAGCAGTGTGAGCTGCTGGTGAAGAGTTTGAAGAGGTGATGAGGACTGTGACAGGAAC	603
DB	344	AAGCAGTGTGAGCTGCTGGTGAAGAGTTTGAAGAGGTGATGAGGACTGTGACAGGAAC	403
QY	604	CACGAGGAGGAGACCTGACCTGAATTCCTCTGTGCAACACGCTGCAAGGGAAGGAC	663
DB	404	CACGAGGAGGAGACCTGACCTGAATTCCTCTGTGCAACACGCTGCAAGGGAAGGAC	463
QY	664	ACGAGTTCCTTAGCAGAGCGGTGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGG	723
DB	464	ACGAGTTCCTTAGCAGAGCGGTGTCTGGCAAGAGGGGGACATAGCCTCCCTGGGAGGG	523
QY	724	AAGAAATCCAGAAAGCGGACGCGAGTCAAGGGCTCCTCAGTGGCAGCAGCAGCAG	783
DB	524	AAGAAATCCAGAAAGCGGACGCGAGTCAAGGGCTCCTCAGTGGCAGCAGCAGCAG	583
QY	784	AGGAGGAACTGGGGGGCTGGGGAGGATGCAACCGCGAGGAGGAGGAGGTGTGCAG	843
DB	584	AGGAGGAACTGGGGGGCTGGGGAGGATGCAACCGCGAGGAGGAGGAGGTGTGCAG	643
QY	844	AAGGATCGCCCTCCCAACACAGCCCTCGATGAGCTGTGAGCCACGCTTAGTGTCTTT	903
DB	644	AAGGATCGCCCTCCCAACACAGCCCTCGATGAGCTGTGAGCCACGCTTAGTGTCTTT	703
QY	904	GAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACGAGCGAGCGAGCGCAG-CTCC	962
DB	704	GAATCAAGACCCCTGACTTCAGAGCTTGGGACACGACGAGCGAGCGAGCGAGCGCCTCC	763
QY	963	AGCAAGGACAGCTGTGTCACGATCAGTCTCCTCCCTTGGCTGTGCCCCCTTTCCTTCC	1022

DB	764	AGCAAGACAGCTGCTGTCCAGCATCAGGTCTCTCTCTTGGCTGTGCCCTTTCCTTCC	823
QY	1023	CTTTGAACAACAGCAAGAGGTGGAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGA	1082
DB	824	CTTTGAACAACAGCAAGAGGTGGAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGA	883
QY	1083	AGAGGAGGAGGAGCAGAAGGACGCTCTTTTCTACAGATCCCTCTCAGAGCTCCGGGG	1142
DB	884	AGAGGAGGAGGAGCAGAAGGACGCTCTTTTCTACAGATCCCTCTCAGAGCTCCGGGG	943
QY	1143	TCACCCAGCATCCCAAGGCTGAGATCCAGGCTCTTACATGGAAGCTGAAGACATGAG	1202
DB	944	TCACCCAGCATCCCAAGGCTGAGATCCAGGCTCTTACATGGAAGCTGAAGACATGAG	1003
QY	1203	GCACATAGATGCTCACAGCGCCCTTACGCGAGGAAGGACTCCGCTGACGCTCAGCA	1262
DB	1004	GCACATAGATGCTCACAGCGCCCTTACGCGAGGAAGGACTCCGCTGACGCTCAGCA	1063
QY	1263	GCCAGGCTGCTCTTCTTCCACCAAGCAATCTTCTGCTGCTCTTGTGCGATGGTA	1322
DB	1064	GCCAGGCTGCTCTTCTTCCACCAAGCAATCTTCTGCTGCTCTTGTGCGATGGTA	1123
QY	1323	AATTGAGAACTTCCAGGACAAACTCGGGTGTGGCACAAAGGGGCTGGAACCCAGAGCA	1382
DB	1124	AATTGAGAACTTCCAGGACAAACTCGGGTGTGGCACAAAGGGGCTGGAACCCAGAGCA	1183
QY	1383	GAGCAGCCAGAGACTGCAAGAGGACCTGACCTAAACCCCTTGAAAGGCAATCTG	1442
DB	1184	GAGCAGCCAGAGACTGCAAGAGGAGGACCTGACCTAAACCCCTTGAAAGGCAATCTG	1243
QY	1443	CAGTTCCTGCTCCACCCACTCTCTGAGGACGCTCATGCTCTGCCAGCCCTTCTCC	1502
DB	1244	CAGTTCCTGCTCCACCCACTCTCTGAGGACGCTCATGCTCTGCCAGCCCTTCTCC	1303
QY	1503	CAGGGCTACCAAGATAAACCTTTTGGCTTTTGGTTCCTGGTTCCTGGTTCCTATCAGC	1562
DB	1304	CAGGGCTACCAAGATAAACCTTTTGGCTTTTGGTTCCTGGTTCCTATCAGC	1363
QY	1563	CTCCAGAGTGTCCCTCATCGATCTTTTGGCTTTTGGTTCCTGGTTCCTATCAGC	1622
DB	1364	CTCCAGAGTGTCCCTCATCGATCTTTTGGCTTTTGGTTCCTGGTTCCTATCAGC	1423
QY	1623	GGCCATCACCATCATTTGAGGCTTAACTGTGAGTACTAGGAGTGTCTGGAGCGCCG	1682
DB	1424	GGCCATCACCATCATTTGAGGCTTAACTGTGAGTACTAGGAGTGTCTGGAGCGCCG	1483
QY	1683	GGGTGGTGGGGTAATCACTCACTGCTCTAGCCTTCTTAACTGAGCCCTTAAAT	1742
DB	1484	GGGTGGTGGGGTAATCACTCACTGCTCTAGCCTTCTTAACTGAGCCCTTAAAT	1543
QY	1743	ACAGTTCCTTCTGTGGTGGTGGTCCACGCGCCCAACACACACACATAAAATTAATTCG	1802
DB	1544	ACAGTTCCTTCTGTGGTGGTGGTCCACGCGCCCAACACACACACATAAAATTAATTCG	1603
QY	1803	ATGCTGTTTCAATACTGT 1820	
DB	1604	ATGCTGTTTCAATACTGT 1621	

RESULT 7

AK082749	1376 bp	mRNA	linear	HTC 03-APR-2004
LOCUS				
DEFINITION	Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone: C330002N05 product: 1600025D17RIK PROTEIN (PUTATIVE RETINOIC ACID-REGULATED PROTEIN) (RIKEN CDNA 1600025D17 GENE), full insert sequence.			
ACCESSION	AK082749			
VERSION	AK082749.1	GI:26349918		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

id and NIH Registry designation is BQ01. Positive for GATYA4, MixL1, MixL1, HNF4alpha expression; negative for AFP expression. Passage number 40. cDNA primed using oligo-dT primer: 5'-pGACTGTTCATGTCGACGCGCCCTT)25-3' and cloned into the EcoRV/NotI sites of pT7-1. This primary library is non-normalized (unnormalized) primary library is NIH_MGC_259). It was constructed by Express Genomics (Frederick, MD). Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Note: this is a Mammalian Gene Collection library."

ORIGIN

Query Match	41.7%	Score 766;	DB 8;	Length 766;
Best Local Similarity	100.0%;	Prod. No. 5.7e-301;		
Matches 766;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	36	CTCTGGGTCGCGCTGGGCCATGGAGTC	CATCTCTGAGCTCGCGCCCGCTCTTTATT	95
Db	1	CTCTGGGTCGCGCTGGGCCATGGAGTC	CATCTCTGAGCTCGCGCCCGCTCTTTATT	60
Qy	96	TCCTTTGCTGCTGCTTCGCTTCGCTCT	CTCTTCCTCCCGGAAAGCTAGGCGCGAGTCC	155
Db	61	TCCTTTGCTGCTGCTTCGCTTCGCTCT	CTCTTCCTCCCGGAAAGCTAGGCGCGAGTCC	120
Qy	156	CGCCGGGGCTGAGGAGACCGACTCGGT	CGCATTCGCCAGCAAAATGCGAAGTGCGCAAGTA	215
Db	121	CGCCGGGGCTGAGGAGACCGACTCGGT	CGCATTCGCCAGCAAAATGCGAAGTGCGCAAGTA	180
Qy	216	TGTTGCTGTGAGCTGAAGTCGGCTCTT	TGAGGAAACGGGAAGACCAAGGAAGTGATTTGA	275
Db	181	TGTTGCTGTGAGCTGAAGTCGGCTCTT	TGAGGAAACGGGAAGACCAAGGAAGTGATTTGA	240
Qy	276	CACCGGCTATGGCATCTCTGGAACGGG	CTCTGGAGTCAAGTACACCAAGTCGGACTT	335
Db	241	CACCGGCTATGGCATCTCTGGAACGGG	CTCTGGAGTCAAGTACACCAAGTCGGACTT	300
Qy	336	ACGGTTAATTCGAAGTCACCTGAGACA	TCATTTGCAAGAGGCTCTGGAGTACAGCTGSCAA	395
Db	301	ACGGTTAATTCGAAGTCACCTGAGACA	TCATTTGCAAGAGGCTCTGGAGTACAGCTGSCAA	360
Qy	396	GGAGAGGACTGGGACCAACCGGTTTG	CCAAAGGATATGTCGAGACCTTTGAGACGCTGCA	455
Db	361	GGAGAGGACTGGGACCAACCGGTTTG	CCAAAGGATATGTCGAGACCTTTGAGACGCTGCA	420
Qy	456	CAACTAGTCCAAAGGGGTCAAGTGGT	GATGATATCCCTATGAGCTGTGGAACGA	515
Db	421	CAACTAGTCCAAAGGGGTCAAGTGGT	GATGATATCCCTATGAGCTGTGGAACGA	480
Qy	516	GACCTCAGCAGAGTGGCTGACCTCA	GAAGCAGTGTGACGTGCTGGTGGAAAGTCTTGA	575
Db	481	GACCTCAGCAGAGTGGCTGACCTCA	GAAGCAGTGTGACGTGCTGGTGGAAAGTCTTGA	540
Qy	576	AGAGGTGATTGAGGACTGGTACAGGAA	CCACAGGAGGAAAGCTGACTGAATTCCTCTG	635
Db	541	AGAGGTGATTGAGGACTGGTACAGGAA	CCACAGGAGGAAAGCTGACTGAATTCCTCTG	600
Qy	636	TGCCAACACGTCGTAAGGGNAAGGA	CACGAGTTGCCCTAGCAGACGGTGTCTGGCAA	695
Db	601	TGCCAACACGTCGTAAGGGNAAGGA	CACGAGTTGCCCTAGCAGACGGTGTCTGGCAA	660
Qy	696	GAAAGGGGACATAGCCTCCCTGGG	AGGGAAGAAATCCAAAGAAAGCCAGCGAGTCAA	755
Db	661	GAAAGGGGACATAGCCTCCCTGGG	AGGGAAGAAATCCAAAGAAAGCCAGCGAGTCAA	720
Qy	756	GGGCTCTCCAGTGGCAGCAACAGCA	GAGGAAGGAACTGGGGGGC 801	
Db	721	GGGCTCTCCAGTGGCAGCAACAGCA	GAGGAAGGAACTGGGGGGC 766	

RESULT 10
BG974240

LOCUS	BC974240	766 bp	mRNA	linear	EST 12-JUN-2000
DEFINITION	602843972F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979944 5', mRNA sequence.				
ACCESSION	BC974240				
VERSION	BC974240.1	GI:14361877			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ ;				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cqabbs-t@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10978 row: o column: 17 High quality sequence stop: 761.				
FEATURES	Location/Qualifiers				
source	1..766 /organism="Mus musculus" /mol_type="mRNA" /strain="NMRI" /db_xref="taxon:10090" /clone="IMAGE:4979944" /tissue_type="tumor, gross tissue" /dev_stage="5 months" /lab_host="DH10B" /clone_lib="NCI_CGAP Mam4" /notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sall; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."				
ORIGIN	41.3%; Score 760; DB 2; Length 766; Query Match 100.0%; Pred. No. 1.6e-298; Indels 0; Gaps 0; Matches 760; Conservative 0; Mismatches 0;				
QY	658 AAGGACACGAGTTGCTTAGCAGACGCGTGTCTGGCAAGAGGGGGACATAGCCTCCCTG 717				
Db	1 AAGGACACGAGTTGCTTAGCAGACGCGTGTCTGGCAAGAGGGGGACATAGCCTCCCTG 60				
QY	718 GGAGGGAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCTCCAGTGGCAGCAGC 777				
Db	61 GGAGGGAAGAAATCCAAAGAAAGCGCAGCGGAGTCAAGGGCTCTCTCCAGTGGCAGCAGC 120				
QY	778 AAGCAGAGGAAGAACTGGGGGGCGCTGGGGGAGATGCCAACGCCGAGGAGGAGGAGGT 837				
Db	121 AAGCAGAGGAAGAACTGGGGGGCGCTGGGGGAGATGCCAACGCCGAGGAGGAGGAGGT 180				
QY	838 GTGCAGAAAGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCACGACTTAGT 897				
Db	181 GTGCAGAAAGCATCGCCCTCCACACAGCCCCCTGATGAGCTGTGAGCCACGACTTAGT 240				
QY	898 GTCCTTGAATCAAGACCCCTGACTTCAAGAGCTTTGGGACACGCACAGCGCAGCGCAGCGCA 957				
Db	241 GTCCTTGAATCAAGACCCCTGACTTCAAGAGCTTTGGGACACGCACAGCGCAGCGCAGCGCA 300				
QY	958 GCTCCAGCAGGACAGCTGCTGTTCAGCATCAGGTCTCTCCCTTGCGCTGTGCCCTTTC 1017				
Db	301 GCTCCAGCAGGACAGCTGCTGTTCAGCATCAGGTCTCTCCCTTGCGCTGTGCCCTTTC 360				

Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5 . Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC Library."

ORIGIN

Query Match 36.9%; Score 679; DB 6; Length 903;
Best Local Similarity 100.0%; Pred. No. 1.5e-265;

1018	Qy	CTTCCCTTGAA	CAACACAG	CAGAGGTG	GAAGGATCTCGGGT	GTCTGGGAG	ACGGCACCCCAA	1077
361	Db	CTTCCCTTGAA	CAACACAG	CAGAGGTG	GAAGGATCTCGGGT	GTCTGGGAG	ACGGCACCCCAA	420
1078	Qy	AGGGAAGAG	GAGGAGG	CAGAGG	CAGCTCTCTTTCT	ACACAGTCCCCCT	CACGAGCTC	1137
421	Db	AGGGAAGAG	GAGGAGG	CAGAGG	CAGCTCTCTTTCT	ACACAGTCCCCCT	CACGAGCTC	480
1138	Qy	CGGGGTCCAC	CCACAG	CATCCC	CAGGCTGAG	ATCCAGGCTCTG	CATGGAAGCTGGAAGAGC	1197
481	Db	CGGGGTCCAC	CCACAG	CATCCC	CAGGCTGAG	ATCCAGGCTCTG	CATGGAAGCTGGAAGAGC	540
1198	Qy	ATGAGGCACA	TAAATAG	TGCTCAC	CAGCGCCCCCTT	CAGCCAGGAAG	AGATCCGCTGCAGCCT	1257
541	Db	ATGAGGCACA	TAAATAG	TGCTCAC	CAGCGCCCCCTT	CAGCCAGGAAG	AGATCCGCTGCAGCCT	600
1258	Qy	CAGCAGCCAG	CGCTGCTCTT	CTCCAC	CAAGCATTTCTT	CTCTGCTGGT	CTCTGTGCGGA	1317
601	Db	CAGCAGCCAG	CGCTGCTCTT	CTCCAC	CAAGCATTTCTT	CTCTGCTGGT	CTCTGTGCGGA	660
1318	Qy	TGTTAAATT	TCGAAACTT	CCAGGACAA	ACTCGGGTGTGG	CAAAAGGGG	CTGACGCCAG	1377
661	Db	TGTTAAATT	TCGAAACTT	CCAGGACAA	ACTCGGGTGTGG	CAAAAGGGG	CTGACGCCAG	720
1378	Qy	AGCCAGAGC	CAACCC	CAGAGACTG	CAGAGGGG	CACTTGAC	1417	
721	Db	AGCCAGAGC	CAACCC	CAGAGACTG	CAGAGGGG	CACTTGAC	760	

RESULT 11	CB320984	903 bp	mRNA	linear	EST 04-MAR-2003
LOCUS	AGENCOURT 12238336	NIH MGC_136	Mus musculus	cDNA clone	
DEFINITION	IMAGE:30288696	5', mRNA	sequence.		
ACCESSION	CB320984				
VERSION	CB320984.1	GI:28845219			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;				
	Sciuroroethalia; Murioidea; Muridae; Murinae; Mus.				
REFERENCE	NIH-MGC	http://mgc.nci.nih.gov/.			
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)				
TITLE	Unpublished (1999)				
JOURNAL	Contact: Robert Strausberg, Ph.D.				
COMMENT	Email: cgaabs-r@mail.nih.gov				
	Tissue Procurement: Dr. David Rowe				
	CDNA Library Preparation: Invitrogen Corp				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: NDAM326	row: j	column: 01		
	High quality	sequence	stop: 611.		

FEATURES

RESULT 12	
BQ922065	
LOCUS	926 bp mRNA linear EST 20-AUG-2002
DEFINITION	AGENCOURT_8932400 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6468664
	5' , mRNA sequence.
ACCESSION	BQ922065
VERSION	BQ922065.1 GI:22337096
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
	Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 926)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: The Cepko Laboratory
Tissue Procurement: The I.M.A.G.E. Consortium (LLNL)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM13996 row: i column: 17
High quality sequence stop: 632.
Location/Qualifiers
1. .926
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:648664"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: Not I; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life technologies. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 36.1%; Score 663; DB 5; Length 956;
Best Local Similarity 100.0%; Pred. No. 5.1e-259; Indels 0; Gaps 0;
Matches 663; Conservative 0; Mismatches 0;
Qy 541 AAGAACGAGTGTGACGTCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGG 600
Db 1 AAGAACGAGTGTGACGTCTGGTGAAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGG 60
Qy 601 AACACACAGGAGGAAGACTGACTGAATTCCTGTGTGCAACACCGAGTCTCTGAAGGGAAG 660
Db 61 AACACACAGGAGGAAGACTGACTGAATTCCTGTGTGCAACACCGAGTCTCTGAAGGGAAG 120
Qy 661 GACACGAGTGTGCTAGCAGAGCGGTGTCTGCAAGAGAGGGGACATAGCTCCTCTGGGA 720
Db 121 GACACGAGTGTGCTAGCAGAGCGGTGTCTGCAAGAGAGGGGACATAGCTCCTCTGGGA 180
Qy 721 GGGAGAGAAATCCAGAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 780
Db 181 GGGAGAGAAATCCAGAGAGAGCGCAGCGGAGTCAAGGGCTCTCCAGTGGCAGCAGCAAG 240
Qy 781 CAGAGGAAGGAATCTGGGGGGCTTGGGGAGAGATGCCAAGCGGAGGAGGAGGGGTGTG 840
Db 241 CAGAGGAAGGAATCTGGGGGGCTTGGGGAGAGATGCCAAGCGGAGGAGGAGGGGTGTG 300
Qy 841 CAGAGGAGATCGCCCTCCACACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 900
Db 301 CAGAGGAGATCGCCCTCCACACAGCCCCCTGTATGAGCTGTGAGCCAGCTTAGTGTG 360
Qy 901 CTTGAATCAAGACCCCTGACTTTACAGCTTTGGGACACGACAGCGCAGCGCAGCT 960
Db 361 CTTGAATCAAGACCCCTGACTTTACAGCTTTGGGACACGACAGCGCAGCGCAGCT 420
Qy 961 CCAGCAAGGACAGCTGTGTCCAGATCAGAGTCTCTCTCTGGGTGTGCCCCCTTTTCCTT 1020
Db 421 CCAGCAAGGACAGCTGTGTCCAGATCAGAGTCTCTCTCTGGGTGTGCCCCCTTTTCCTT 480
Qy 1021 CCTTTGAACACAGCAAGAGGTGGAAGATCTGGGGTCTGGGACACGGCACCCCAAAGG 1080
Db 481 CCTTTGAACACAGCAAGAGGTGGAAGATCTGGGGTCTGGGACACGGCACCCCAAAGG 540
Qy 1081 GAAGAGGAGGAGGAGCAGAGGAGCAGCTCTCTTTTACACAGTCCCTTCACGAGTCCGG 1140
Db 541 GAAGAGGAGGAGGAGCAGAGGAGCAGCTCTCTTTTACACAGTCCCTTCACGAGTCCGG 600

Qy 1141 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATCGAAGCTGAAGGCAATG 1200
Db 601 GGTCCACCCAGCATCCCGAGGCTGAGATCCAGGCTCCTGACATCGAAGCTGAAGGCAATG 660
Qy 1201 AGG 1203
Db 661 AGG 663
RESULT 13
CF553710
LOCUS
DEFINITION AGENCOURT 15614073 Organ of Corti Mus musculus cDNA clone
IMAGE:30519414 5', mRNA sequence.
ACCESSION CF553710
VERSION CF553710
KEYWORDS EST.
SOURCE CF553710.1 GI:34890544
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 892)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Bechara Kachar and Celine Pompeia, NIDCD
cDNA Library Preparation: Bechara Kachar and Celine Pompeia, NIDCD
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM590 row: O column: 07
High quality sequence stop: 579.
Location/Qualifiers
1. .892
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30519414"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Organ of Corti"
/note="Organ: Organ of Corti/Ear; Vector: pBluescriptSK-; Site 1: EcoRI; Site 2: XhoI; The organ of Corti (OC) was fine dissected from a total of 386 OC as follows: 102 samples from post-natal (P) day 5; 72 from P6; 60 from P7; 46 from P8; 18 from P9; 20 from P10; 14 from P12 and 24 from P13. Total RNA was extracted using the micro Fasttrack kit (catalog # K1593-02; Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Reverse transcription and library construction were carried out with the Uni-Zap XR vector kit (catalog # 237211, Stratagene) and Uni-Zap XR Gigapack III Gold Cloning kit (catalog # 237612), both from Stratagene (La Jolla, CA, USA), according to manufacturer's instructions. The frequency distribution of the library is as follows: 72% of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and 0.1% 51-150. As to gene function, 45% of genes are present in GenBank and have known function; 23% have hits in GenBank, but do not have assigned function; 12% are uncharacterized ESTs and 20% are unidentified. Library created in the laboratory of M. Brownstein (NIMH, NIH). A complete library description can be found at ."
FEATURES
source
ORIGIN
Query Match 36.0%; Score 662; DB 6; Length 892;

Best Local Similarity 100.0%; Pred. No. 1.3e-258;		Matches 662; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	49	TGGGCGATGAGTCCATGCTGAGCTCGCGCCCGCTGCTCTTATTTCTTTGCTGCTG	108
Db	9	TGGGCGATGAGTCCATGCTGAGCTCGCGCCCGCTGCTCTTATTTCTTTGCTGCTG	68
Qy	109	CTGCTTCGCGTCTGCTCTTCTGCGCCGGAAGCTAGGCGCGAGTCCCGCGGGCTGAG	168
Db	69	CTGCTTCGCGTCTGCTCTTCTGCGCCGGAAGCTAGGCGCGAGTCCCGCGGGCTGAG	128
Qy	169	GAGACCGACTGGTGGATGCGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAG	228
Db	129	GAGACCGACTGGTGGATGCGCCAGCAAAATGCGAAGTGTGCAAGTATGTTGCTGTGGAG	188
Qy	229	CTGAAGTCTGGCTTTTGGAGAAACGGGAAGACCAAGGAAGTGAATGACACCGGCTATGGC	288
Db	189	CTGAAGTCTGGCTTTTGGAGAAACGGGAAGACCAAGGAAGTGAATGACACCGGCTATGGC	248
Qy	289	ATCTCGGACGGGAAGGCTCTGAGTCAAGTACACAAAGTCGAGCTTACGGTTAATTGAA	348
Db	249	ATCTCGGACGGGAAGGCTCTGAGTCAAGTACACAAAGTCGAGCTTACGGTTAATTGAA	308
Qy	349	GTCAGTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCTGCACAGGAGGACTGCG	408
Db	309	GTCAGTGAGACCATTTGCAAGAGGCTTCTGGACTACAGCTGCACAGGAGGACTGCG	368
Qy	409	AGCAACCGGTTTCCCAAGGATGTGTCGAGACCTTTTGAGACCTGCACAACTAGTCCAC	468
Db	369	AGCAACCGGTTTCCCAAGGATGTGTCGAGACCTTTTGAGACCTGCACAACTAGTCCAC	428
Qy	469	AAAGGGGTCAAGTGGTGAATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAG	528
Db	429	AAAGGGGTCAAGTGGTGAATATCCCTATGAGCTGTGGAACGAGACCTCAGCAGAG	488
Qy	529	GTGGCTGACCTCAAGAACAGTGTGACCTGTGGTGGAGAGTTTGAAGAGTGAATTGAG	588
Db	489	GTGGCTGACCTCAAGAACAGTGTGACCTGTGGTGGAGAGTTTGAAGAGTGAATTGAG	548
Qy	589	GACTGGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACACCACTG	648
Db	549	GACTGGTACAGGAACCAACAGGAGGAACCTGACTGAATTCCTCTGTGCCAACACCACTG	608
Qy	649	CTGAAGGAAAGGACACAGGTTGCTCTAGCAGCGGTGTCTGGCAAGAGGGGACATA	708
Db	609	CTGAAGGAAAGGACACAGGTTGCTCTAGCAGCGGTGTCTGGCAAGAGGGGACATA	668
Qy	709	GC 710	
Db	669	GC 670	

RESULT 14
CB203728
LOCUS
DEFINITION
IMAGE:30142422 5', mRNA sequence.
ACCESSION
CB203728
VERSION
CB203728.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 937)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM0051 row: k column: 07
High quality sequence stop: 609.
Location/Qualifiers
1. 937
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30142422"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_135"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized full-length enriched library from pooled mouse
embryonic limb, maxilla and mandible, day 12.5, 13.5,
14.5, and 15.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dr. cDNA
enrichment: >1k bp. Average insert size 1.6k bp.
Normalization (Cot value): 7.5 kb. Priming sequence:
5'GACTAGTCTAGATCGAGCGCGCC(T)3' Tissue contributed
by, David Rowe. Library constructed by ResGen, Invitrogen
Corp."

ORIGIN

Query Match		35.7%;		Score 656;		DB 6;		Length 937;	
Best Local Similarity		99.9%;		Pred. No. 3.6e-256;					
Matches		706;		Conservative		0;		Mismatches 1; Indels 0; Gaps 0;	
Qy	559	CTGTGTGAAGAGCTTTGAAGAGGTGATTGAGGACTGTGTACAGGAACCAACAGGAGGAAGAC	618	Db	1	CTGTGTGAAGAGCTTTGAAGAGGTGATTGAGGACTGTGTACAGGAACCAACAGGAGGAAGAC	60		
Qy	619	CTGACTGAATTCCTCTGTGCCAAACACCGTGTGAAGGGAAGGACACGAGTTGCCTAGCA	678	Db	61	CTGACTGAATTCCTCTGTGCCAAACACCGTGTGAAGGGAAGGACACGAGTTGCCTAGCA	120		
Qy	679	GAGCGGTGTCGTGGCAAGAGGGGACATAGCTCTCTGGGAGGGAAGAAATCCAAAGAG	738	Db	121	GAGCGGTGTCGTGGCAAGAGGGGACATAGCTCTCTGGGAGGGAAGAAATCCAAAGAG	180		
Qy	739	AAGCCGACGGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAGGAAGAACTGGGG	798	Db	181	AAGCCGACGGAGTCAAGGGCTCTCTCAGTGGCAGCAGCAAGCAGAGGAGGAAGAACTGGGG	240		
Qy	799	GGCTTGGGGAGGATGCCAACCGAGGAGGAGGGGTGTGCAGAGGCATCGCCCCTC	858	Db	241	GGCTTGGGGAGGATGCCAACCGAGGAGGAGGGGTGTGCAGAGGCATCGCCCCTC	300		
Qy	859	CCACACAGCCCCCTGATGAGCTGTAGCCCGCAGCTTAGTGTCTTGAATCAAGACCCCTG	918	Db	301	CCACACAGCCCCCTGATGAGCTGTAGCCCGCAGCTTAGTGTCTTGAATCAAGACCCCTG	360		
Qy	919	ACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCAGCAGGACAGCTGCT	978	Db	361	ACTTCAGAGCTTGGGACACGACAGCGCAGCGCAGCTCAGCAGGACAGCTGCT	420		
Qy	979	GTCCAGCATCAGTCTCTCTCTGCTGCTGCTGCTTCTTCTTCTTGAACACGCAAG	1038	Db	421	GTCCAGCATCAGTCTCTCTCTGCTGCTGCTTCTTCTTCTTGAACACGCAAG	480		
Qy	1039	AGGTGGAAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGAAGAGGAGGAGGACAG	1098	Db	481	AGGTGGAAGGATCTGGGGTGTCTGGGAGACGGCACCCCAAGGGAAGAGGAGGAGGACAG	540		
Qy	1099	AAGGCAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGGGTCTCCACCGAGATCCCC	1158	Db	541	AAGGCAGCTCTCTTTCTACACAGTCCCTCTCAGAGCTCCGGGTCTCCACCGAGATCCCC	600		
Qy	1159	AGGCTGAGATCCAGGCTCTTGATGGAAGCTGAAGAGCATGAGGACATAGATGCTCA	1218						

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 04:07:46 ; Search time 355 Seconds
(without alignments)
9208.270 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgc.....taaaaaaaaaaaaaaaaa 1839

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 1303057 seqs, 888780828 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PC/US COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	2.7	1201	3	US-09-023-655-140
2	36	2.0	1534	3	US-09-023-655-274
3	33	1.8	1658	3	US-09-482-273-23
4	26	1.4	601	3	US-09-949-016-132624
5	26	1.4	450395	3	US-09-949-016-15473
6	25	1.4	601	3	US-09-949-016-23380
7	25	1.4	601	3	US-09-949-016-141221
8	25	1.4	601	3	US-09-949-016-151342
9	25	1.4	601	3	US-09-949-016-151343
10	25	1.4	601	3	US-09-949-016-151405
11	25	1.4	601	3	US-09-949-016-151406
12	25	1.4	1589	3	US-09-949-016-3999
13	25	1.4	9087	3	US-09-949-016-15158
14	25	1.4	11211	3	US-09-949-016-13251
15	25	1.4	18601	3	US-09-949-016-15629
16	25	1.4	33753	3	US-09-949-016-15741
17	25	1.4	33756	3	US-09-949-016-12006
18	25	1.4	102520	3	US-09-949-016-17367
19	25	1.4	102526	3	US-09-949-016-12448
20	25	1.4	118382	3	US-09-949-016-15996
21	25	1.4	118382	3	US-09-949-016-15997
22	24	1.3	321	2	US-08-171-385-22
23	24	1.3	321	3	US-08-361-441B-22
24	24	1.3	321	3	US-08-171-385-27

C 25	24	1.3	340	3	US-08-361-441B-27	Sequence 27, Appl
C 26	24	1.3	345	3	US-08-171-385-20	Sequence 20, Appl
C 27	24	1.3	345	3	US-08-361-441B-20	Sequence 20, Appl
C 28	24	1.3	413	3	US-09-014-969-5	Sequence 5, Appl
C 29	24	1.3	1073	3	US-09-205-258-127	Sequence 127, App
C 30	24	1.3	1073	3	US-10-004-860-127	Sequence 127, App
C 31	24	1.3	2837	2	US-08-993-228-11	Sequence 11, Appl
C 32	24	1.3	3453	2	US-07-813-593-1	Sequence 1, Appl
C 33	24	1.3	3453	2	US-07-977-451-1	Sequence 1, Appl
C 34	24	1.3	3453	2	US-07-946-507-1	Sequence 1, Appl
C 35	24	1.3	3453	2	US-08-252-517-1	Sequence 1, Appl
C 36	24	1.3	3453	2	US-07-906-397A-1	Sequence 1, Appl
C 37	24	1.3	3453	2	US-08-601-891-1	Sequence 1, Appl
C 38	24	1.3	3453	2	US-09-021-324-1	Sequence 1, Appl
C 39	24	1.3	3453	3	US-09-872-136B-1	Sequence 1, Appl
C 40	24	1.3	3453	3	US-09-919-408A-1	Sequence 1, Appl
C 41	24	1.3	3453	6	PCT-US92-02750-1	Sequence 1, Appl
C 42	24	1.3	3453	6	PCT-US92-05401-1	Sequence 1, Appl
C 43	24	1.3	3453	6	PCT-US92-09893-1	Sequence 1, Appl
C 44	24	1.3	11111	3	US-09-949-016-14960	Sequence 14960, A
C 45	24	1.3	48691	3	US-09-949-016-16308	Sequence 16308, A

ALIGNMENTS

RESULT 1

US-09-023-655-140
; Sequence 140, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THPINOB01
; CLONE: 030137

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15473
; LENGTH: 450395
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)-(450395)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15473

Query Match 1.4%; Score 26; DB 3; Length 450395;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1814 TAACGTGTAATAAAAAAAAAAAAAAAAAAAAA 1839
Db 418347 TAACGTGTAATAAAAAAAAAAAAAAAAAAAAA 418372

RESULT 6

US-09-949-016-23380
; Sequence 23380, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23380
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-23380

Query Match 1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAAAAA 1839
Db 222 AACTGTAAAAAAAAAAAAAAAAAAAA 246

RESULT 7

US-09-949-016-141221
; Sequence 141221, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141221
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-141221

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141221
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-141221

Query Match 1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAAAAA 1839
Db 222 AACTGTAAAAAAAAAAAAAAAAAAAA 246

RESULT 8

US-09-949-016-151342
; Sequence 151342, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151342
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151342

Query Match 1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAAAAAAAAAAAAAAAAA 1839
Db 350 AACTGTAAAAAAAAAAAAAAAAAAAA 374

RESULT 9

US-09-949-016-151343
; Sequence 151343, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151343
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151343

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151343
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151343

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 365 AACTGTAAAAA 389

RESULT 10
US-09-949-016-151405
; Sequence 151405, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151405
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151405

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 350 AACTGTAAAAA 374

RESULT 11
US-09-949-016-151406
; Sequence 151406, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 151406
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-151406

Query Match          1.4%; Score 25; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 365 AACTGTAAAAA 389

RESULT 12
US-09-949-016-3999/c
; Sequence 3999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3999
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3999

Query Match          1.4%; Score 25; DB 3; Length 1589;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 1433 AACTGTAAAAA 1409

RESULT 13
US-09-949-016-15158/c
; Sequence 15158, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15158
; LENGTH: 9087
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(9087)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15158

Query Match      1.4%; Score 25; DB 3; Length 9087;
Best Local Similarity 100.0%; Pred.No. 0.93;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 1186 AACTGTAAAAA 1162

RESULT 14
US-09-949-016-13251
; Sequence 13251, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13251
; LENGTH: 11211
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(11211)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13251
```

```
Query Match      1.4%; Score 25; DB 3; Length 11211;
Best Local Similarity 100.0%; Pred.No. 0.93;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 11122 AACTGTAAAAA 11146
```

```
RESULT 15
US-09-949-016-15629/c
; Sequence 15629, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15629
; LENGTH: 19601
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15629

Query Match      1.4%; Score 25; DB 3; Length 19601;
Best Local Similarity 100.0%; Pred.No. 0.89;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
Db 18157 AACTGTAAAAA 18133
```

```
Search completed: December 30, 2005, 09:11:04
Job time : 357 secs
```

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	100	5.4	100	7	US-10-430-201-4789	Sequence 4789, App
2	100	5.4	100	7	US-10-430-201-4790	Sequence 4790, App
3	50	2.7	444	7	US-10-424-599-33803	Sequence 33803, A
4	50	2.7	580	3	US-09-777-564-1232	Sequence 1232, App
5	50	2.7	580	5	US-10-015-219-1232	Sequence 1232, App
6	50	2.7	901	7	US-10-262-839-69	Sequence 69, Appl
7	50	2.7	1201	7	US-10-641-643-139	Sequence 140, App
8	50	2.7	1378	6	US-10-126-103-139	Sequence 139, App
9	50	2.7	1378	7	US-10-431-096-139	Sequence 139, App
10	50	2.7	1596	7	US-10-262-839-67	Sequence 67, Appl
11	50	2.7	1675	5	US-10-028-072-135	Sequence 135, App
12	50	2.7	1675	5	US-10-140-808-135	Sequence 135, App
13	50	2.7	1675	5	US-10-121-049-135	Sequence 135, App
14	50	2.7	1675	5	US-10-123-904-135	Sequence 135, App
15	50	2.7	1675	5	US-10-140-470-135	Sequence 135, App
16	50	2.7	1675	5	US-10-175-746-135	Sequence 135, App
17	50	2.7	1675	5	US-10-176-918-135	Sequence 135, App
18	50	2.7	1675	5	US-10-176-921-135	Sequence 135, App
19	50	2.7	1675	5	US-10-137-865-135	Sequence 135, App
20	50	2.7	1675	5	US-10-140-474-135	Sequence 135, App
21	50	2.7	1675	5	US-10-142-431-135	Sequence 135, App
22	50	2.7	1675	5	US-10-143-114-135	Sequence 135, App
23	50	2.7	1675	5	US-10-142-419-135	Sequence 135, App


```

; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: <Unknown>
;   FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Zeller, Karen J.
;   REGISTRATION NUMBER: 37,071
;   REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 855-0555
;   TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 140:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1201 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: THPINOB01
;   CLONE: 030137
; SEQUENCE DESCRIPTION: SEQ ID NO: 140 :
US-10-641-643-140

      Query Match          2.7%; Score 50; DB 7; Length 1201;
      Best Local Similarity 100.0%; Pred. No. 3.2e-14;
      Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      586 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTG 635
      |||||||
DB      421 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTG 470

RESULT 8
US-10-126-103-139
; Sequence 139, Application US/10126103
; Publication No. US20030224486A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-
;   FILE REFERENCE: D0108.np
;   CURRENT APPLICATION NUMBER: US/10/126,103
;   CURRENT FILING DATE: 2002-04-19
;   PRIOR APPLICATION NUMBER: US 60/284,962
;   PRIOR FILING DATE: 2001-04-19
;   PRIOR APPLICATION NUMBER: US 60/286,645
;   PRIOR FILING DATE: 2001-04-26
;   PRIOR APPLICATION NUMBER: US 60/346,986
;   PRIOR FILING DATE: 2002-01-09
;   NUMBER OF SEQ ID NOS: 284
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 139
;   LENGTH: 1378
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-126-103-139

      Query Match          2.7%; Score 50; DB 6; Length 1378;
      Best Local Similarity 100.0%; Pred. No. 3.1e-14;
      Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      586 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTG 635
      |||||||
DB      602 GAGGACTGGTACAGGAACCAACAGGAGGAAGACCTGACTGAATTCCTCTG 651

RESULT 9
US-10-431-096-139
; Sequence 139, Application US/10431096
; Publication No. US2004008696A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-
;   FILE OF INVENTION: PATHWAY

```

```

; FILE REFERENCE: D0108A CIP
; CURRENT APPLICATION NUMBER: US/10/431,096
; PRIOR FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: US 60/284,962
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 10/126,103
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/286,645
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/346,986
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 139
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-431-096-139

Query Match      2.7%; Score 50; DB 7; Length 1378;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 602 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAATTCCTCTG 651

RESULT 10
US-10-262-839-67
; Sequence 67, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Smithson, Glennnda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16

```

```

; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 67
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (388)..(549)
US-10-262-839-67

Query Match      2.7%; Score 50; DB 7; Length 1596;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAATTCCTCTG 635
Db 795 GAGGACTGGTACAGGAACACACAGGAGGAGACCTGACTGAATTCCTCTG 844

RESULT 11
US-10-028-072-135
; Sequence 135, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/045911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18

```

1	PRIOR APPLICATION NUMBER: 60/059352	2	PRIOR FILING DATE: 1998-03-12
3	PRIOR FILING DATE: 1997-09-19	4	PRIOR APPLICATION NUMBER: 60/078910
5	PRIOR APPLICATION NUMBER: 60/059588	6	PRIOR FILING DATE: 1998-03-20
7	PRIOR FILING DATE: 1997-09-19	8	PRIOR APPLICATION NUMBER: 60/079294
9	PRIOR APPLICATION NUMBER: 60/059836	10	PRIOR FILING DATE: 1998-03-25
11	PRIOR FILING DATE: 1997-09-24	12	PRIOR APPLICATION NUMBER: 60/079663
13	PRIOR APPLICATION NUMBER: 60/062250	14	PRIOR FILING DATE: 1998-02-27
15	PRIOR FILING DATE: 1997-10-17	16	PRIOR APPLICATION NUMBER: 60/079728
17	PRIOR APPLICATION NUMBER: 60/062285	18	PRIOR FILING DATE: 1998-03-27
19	PRIOR FILING DATE: 1997-10-17	20	PRIOR APPLICATION NUMBER: 60/080165
21	PRIOR APPLICATION NUMBER: 60/062287	22	PRIOR FILING DATE: 1998-03-31
23	PRIOR FILING DATE: 1997-10-17	24	PRIOR APPLICATION NUMBER: 60/081203
25	PRIOR APPLICATION NUMBER: 60/062814	26	PRIOR FILING DATE: 1998-04-09
27	PRIOR FILING DATE: 1997-10-24	28	PRIOR APPLICATION NUMBER: 60/081229
29	PRIOR APPLICATION NUMBER: 60/062816	30	PRIOR FILING DATE: 1998-04-09
31	PRIOR FILING DATE: 1997-10-24	32	PRIOR APPLICATION NUMBER: 60/081695
33	PRIOR APPLICATION NUMBER: 60/063045	34	PRIOR FILING DATE: 1998-04-14
35	PRIOR FILING DATE: 1997-10-24	36	PRIOR APPLICATION NUMBER: 60/081817
37	PRIOR APPLICATION NUMBER: 60/063082	38	PRIOR FILING DATE: 1998-04-15
39	PRIOR FILING DATE: 1997-10-31	40	PRIOR APPLICATION NUMBER: 60/081818
41	PRIOR APPLICATION NUMBER: 60/063127	42	PRIOR FILING DATE: 1998-04-15
43	PRIOR FILING DATE: 1997-10-24	44	PRIOR APPLICATION NUMBER: 60/082999
45	PRIOR APPLICATION NUMBER: 60/063327	46	PRIOR FILING DATE: 1998-04-24
47	PRIOR FILING DATE: 1997-10-27	48	PRIOR APPLICATION NUMBER: 60/083322
49	PRIOR APPLICATION NUMBER: 60/063329	50	PRIOR FILING DATE: 1998-04-28
51	PRIOR FILING DATE: 1997-10-27	52	PRIOR APPLICATION NUMBER: 60/083545
53	PRIOR APPLICATION NUMBER: 60/063550	54	PRIOR FILING DATE: 1998-04-29
55	PRIOR FILING DATE: 1997-10-28	56	PRIOR APPLICATION NUMBER: 60/084600
57	PRIOR APPLICATION NUMBER: 60/063561	58	PRIOR FILING DATE: 1998-05-07
59	PRIOR FILING DATE: 1997-10-28	60	PRIOR APPLICATION NUMBER: 60/084627
61	PRIOR APPLICATION NUMBER: 60/063704	62	PRIOR FILING DATE: 1998-05-07
63	PRIOR FILING DATE: 1997-10-29	64	PRIOR APPLICATION NUMBER: 60/084637
65	PRIOR APPLICATION NUMBER: 60/063733	66	PRIOR FILING DATE: 1998-05-07
67	PRIOR FILING DATE: 1997-10-29	68	PRIOR APPLICATION NUMBER: 60/085149
69	PRIOR APPLICATION NUMBER: 60/063735	70	PRIOR FILING DATE: 1998-05-12
71	PRIOR FILING DATE: 1997-10-29	72	PRIOR APPLICATION NUMBER: 60/085323
73	PRIOR APPLICATION NUMBER: 60/064248	74	PRIOR FILING DATE: 1998-05-13
75	PRIOR FILING DATE: 1997-11-03	76	PRIOR APPLICATION NUMBER: 60/085338
77	PRIOR APPLICATION NUMBER: 60/064809	78	PRIOR FILING DATE: 1998-05-13
79	PRIOR FILING DATE: 1997-11-07	80	PRIOR APPLICATION NUMBER: 60/085339
81	PRIOR APPLICATION NUMBER: 60/065186	82	PRIOR FILING DATE: 1998-05-13
83	PRIOR FILING DATE: 1997-11-12	84	PRIOR APPLICATION NUMBER: 60/085579
85	PRIOR APPLICATION NUMBER: 60/065846	86	PRIOR FILING DATE: 1998-05-15
87	PRIOR FILING DATE: 1997-11-17	88	PRIOR APPLICATION NUMBER: 60/085697
89	PRIOR APPLICATION NUMBER: 60/066364	90	PRIOR FILING DATE: 1998-05-15
91	PRIOR FILING DATE: 1997-11-21	92	PRIOR APPLICATION NUMBER: 60/085704
93	PRIOR APPLICATION NUMBER: 60/066453	94	PRIOR FILING DATE: 1998-05-15
95	PRIOR FILING DATE: 1997-11-24	96	PRIOR APPLICATION NUMBER: 60/086414
97	PRIOR APPLICATION NUMBER: 60/066511	98	PRIOR FILING DATE: 1998-05-22
99	PRIOR FILING DATE: 1997-11-24	100	PRIOR APPLICATION NUMBER: 60/086430
101	PRIOR APPLICATION NUMBER: 60/066770	102	PRIOR FILING DATE: 1998-05-22
103	PRIOR FILING DATE: 1997-11-24	104	PRIOR APPLICATION NUMBER: 60/088730
105	PRIOR APPLICATION NUMBER: 60/069212	106	PRIOR FILING DATE: 1998-06-10
107	PRIOR FILING DATE: 1997-12-11	108	PRIOR APPLICATION NUMBER: 60/088741
109	PRIOR APPLICATION NUMBER: 60/069278	110	PRIOR FILING DATE: 1998-06-10
111	PRIOR FILING DATE: 1997-12-11	112	PRIOR APPLICATION NUMBER: 60/088810
113	PRIOR APPLICATION NUMBER: 60/069334	114	PRIOR FILING DATE: 1998-06-10
115	PRIOR FILING DATE: 1997-12-11	116	PRIOR APPLICATION NUMBER: 60/088858
117	PRIOR APPLICATION NUMBER: 60/069694	118	PRIOR FILING DATE: 19/98-06-11
119	PRIOR FILING DATE: 1997-12-16	120	PRIOR APPLICATION NUMBER: 60/089532
121	PRIOR APPLICATION NUMBER: 60/072320	122	PRIOR FILING DATE: 1998-06-17
123	PRIOR FILING DATE: 1998-01-23	124	PRIOR APPLICATION NUMBER: 60/089907
125	PRIOR APPLICATION NUMBER: 60/073612	126	PRIOR FILING DATE: 1998-06-18
127	PRIOR FILING DATE: 1998-02-04	128	PRIOR APPLICATION NUMBER: 60/089947
129	PRIOR APPLICATION NUMBER: 60/074086	130	PRIOR FILING DATE: 1998-06-19
131	PRIOR FILING DATE: 1998-02-09	132	PRIOR APPLICATION NUMBER: 60/074092
133	PRIOR APPLICATION NUMBER: 60/074092	134	PRIOR FILING DATE: 1998-02-09
135	PRIOR FILING DATE: 1998-02-09	136	PRIOR APPLICATION NUMBER:

Qy 586 GAGGACTGGTACAGGAACCACCAGGAGGAAGACTGTGACTGAATTCCTCTG 635
|||||
Db 868 GAGGACTGGTACAGGAACCACCAGGAGGAAGACTGTGACTGAATTCCTCTG 917

```
Query Match      2.7%; Score 50; DB 5; Length 1675;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 50: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 586 GAGGACTGTTACAGGAACCAACGAGGGAAGACCTGACTGAATTCCTCTG 635
|||||
Db 868 GAGGACTGTTACAGGAACCAACGAGGGAAGACCTGACTGAATTCCTCTG 917

Search completed: December 30, 2005, 09:34:20
Job time : 1383 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 04:41:12 ; Search time 299 Seconds
(without alignments)
3224.269 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggaagcc.....taaaaaaaaaaaaaaa 1839

Scoring table: OLIIGO NUC
Gapop 60.0, Gapext 60.0

Searched: 4172979 seqs, 262114271 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US12_NEW_PUB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US13_NEW_PUB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	2.7	1675	6	US-10-131-826A-135
2	25	1.4	25	7	US-11-121-849-665859
3	25	1.4	25	7	US-11-121-849-665860
4	25	1.4	1612	6	US-10-750-185-3528
5	25	1.4	1833	6	US-10-750-185-38459
6	25	1.4	193789	7	US-11-112-908-55
7	24	1.3	25	7	US-11-121-849-665854
8	24	1.3	14911	7	US-11-112-908-63
9	24	1.3	154548	7	US-11-121-086-33
10	24	1.3	157230	7	US-11-112-908-64
11	24	1.3	158692	7	US-11-121-086-30
12	24	1.3	170508	7	US-11-112-908-62
13	24	1.3	172649	7	US-11-121-086-36
14	24	1.3	173115	7	US-11-112-908-65
15	24	1.3	193363	7	US-11-112-908-32
16	23	1.3	25	7	US-11-121-849-665857
17	23	1.3	946	6	US-10-750-185-37066
18	23	1.3	1120	6	US-10-750-185-29060
19	23	1.3	1382	6	US-10-750-185-39793
20	23	1.3	1445	6	US-10-750-185-29927
21	23	1.3	2352	6	US-10-909-125-835
22	23	1.3	10166	7	US-11-147-606-3
23	23	1.3	172147	7	US-11-112-908-22

c 24	23	1.3	188682	7	US-11-112-908-23
c 25	23	1.3	246960	7	US-11-121-086-8
c 26	22	1.2	201	6	US-10-995-561-18671
c 27	22	1.2	201	6	US-10-995-561-35450
c 28	22	1.2	201	6	US-10-995-561-35535
c 29	22	1.2	201	6	US-10-995-561-43529
c 30	22	1.2	201	6	US-10-995-561-46985
c 31	22	1.2	201	6	US-10-995-561-47049
c 32	22	1.2	201	6	US-10-995-561-50304
c 33	22	1.2	201	6	US-10-995-561-50700
c 34	22	1.2	201	6	US-10-995-561-50702
c 35	22	1.2	201	6	US-10-995-561-50871
c 36	22	1.2	201	6	US-10-995-561-53197
c 37	22	1.2	201	6	US-10-995-561-58387
c 38	22	1.2	201	6	US-10-995-561-80946
c 39	22	1.2	201	6	US-10-995-561-82910
c 40	22	1.2	576	7	US-11-156-909-2
c 41	22	1.2	598	6	US-10-750-185-27071
c 42	22	1.2	600	6	US-10-750-185-21696
c 43	22	1.2	729	6	US-10-750-185-34079
c 44	22	1.2	776	7	US-11-156-909-34
c 45	22	1.2	778	6	US-10-986-501-38

ALIGNMENTS

RESULT 1
US-10-131-826A-135
; Sequence 135, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588

; PRIOR FILING DATE: 1997-09-19
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 135
 ; LENGTH: 1675
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-131-826A-135

Query Match 2.7%; Score 50; DB 6; Length 1675;
 Best Local Similarity 100.0%; Pred. No. 8.3e-11;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 586 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 635
 Db 868 GAGGACTGGTACAGGAAACCCAGGAGGAGACCTGACTGAATTCCTCTG 917

RESULT 2
 US-11-121-849-665859
 ; Sequence 665859, Application US/11121849
 ; Publication No. US20050272080A1
 ; GENERAL INFORMATION:
 ; APPLICANT: John Palma
 ; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
 ; FILE REFERENCE: Microarrays
 ; CURRENT APPLICATION NUMBER: US/11/121,849
 ; CURRENT FILING DATE: 2005-05-03
 ; PRIOR APPLICATION NUMBER: 60/567,949
 ; PRIOR FILING DATE: 2004-05-03
 ; NUMBER OF SEQ ID NOS: 673904
 ; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
 ; SEQ ID NO 665859
 ; LENGTH: 25
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-11-121-849-665859

Query Match 1.4%; Score 25; DB 7; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 TACAGGACCCAGGAGGAGACC 619
 Db 1 TACAGGACCCAGGAGGAGACC 25

RESULT 3
 US-11-121-849-665860
 ; Sequence 665860, Application US/11121849
 ; Publication No. US20050272080A1
 ; GENERAL INFORMATION:
 ; APPLICANT: John Palma
 ; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
 ; FILE REFERENCE: Microarrays
 ; CURRENT APPLICATION NUMBER: US/11/121,849
 ; CURRENT FILING DATE: 2005-05-03
 ; PRIOR APPLICATION NUMBER: 60/567,949
 ; PRIOR FILING DATE: 2004-05-03
 ; NUMBER OF SEQ ID NOS: 673904
 ; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
 ; SEQ ID NO 665860
 ; LENGTH: 25
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-11-121-849-665860

Query Match 1.4%; Score 25; DB 7; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 606 CCAGGAGGAGACCTGACTGAATTC 630
 Db 1 CCAGGAGGAGACCTGACTGAATTC 25
 RESULT 4
 US-10-750-185-35528
 ; Sequence 35528, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 35528
 ; LENGTH: 1612
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866880774774
 US-10-750-185-35528

Query Match 1.4%; Score 25; DB 6; Length 1612;
 Best Local Similarity 100.0%; Pred. No. 0.56;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
 Db 1242 AACTGTAAAAA 1266

RESULT 5
 US-10-750-185-38459/c
 ; Sequence 38459, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 38459
 ; LENGTH: 1833
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866880774774
 US-10-750-185-38459

Query Match 1.4%; Score 25; DB 6; Length 1833;
 Best Local Similarity 100.0%; Pred. No. 0.55;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1815 AACTGTAAAAA 1839
 Db 1789 AACTGTAAAAA 1765

```
RESULT 6
US-11-112-908-55/c
; Sequence 55, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 55
; LENGTH: 193789
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-55

Query Match      1.4%; Score 25; DB 7; Length 193789;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1815 AACTGTAAAAA... 1839
Db 5980 AACTGTAAAAA... 5956

RESULT 7
US-11-121-849-665854
; Sequence 665854, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 665854
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-665854

Query Match      1.3%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 CCCCTATGAGCTGTGGAACGAC 518
Db 1 CCCCTATGAGCTGTGGAACGAC 24

RESULT 8
US-11-112-908-63/c
; Sequence 63, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 63
; LENGTH: 149111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-63

Query Match      1.3%; Score 24; DB 7; Length 149111;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 ACTGTAAAAA... 1839
Db 136117 ACTGTAAAAA... 136094

RESULT 9
US-11-121-086-33
; Sequence 33, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: FOULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 33
; LENGTH: 154548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-33

Query Match      1.3%; Score 24; DB 7; Length 154548;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1816 ACTGTAAAAA... 1839
Db 71640 ACTGTAAAAA... 71663

RESULT 10
US-11-112-908-64/c
; Sequence 64, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 63
; LENGTH: 149111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-63
```

```
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 64
; LENGTH: 157230
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-64
```

```
Query Match 1.3%; Score 24; DB 7; Length 157230;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1816 ACTGTAAAAA 1839
Db 83463 ACTGTAAAAA 83440
```

```
RESULT 11
US-11-121-086-30
; Sequence 30, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 30
; LENGTH: 158692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-30
```

```
Query Match 1.3%; Score 24; DB 7; Length 158692;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1816 ACTGTAAAAA 1839
Db 40153 ACTGTAAAAA 40176
```

```
RESULT 12
US-11-112-908-62/c
; Sequence 62, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
```

```
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 62
; LENGTH: 170508
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-62
```

```
Query Match 1.3%; Score 24; DB 7; Length 170508;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1816 ACTGTAAAAA 1839
Db 67219 ACTGTAAAAA 67196
```

```
RESULT 13
US-11-121-086-36
; Sequence 36, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 172649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-36
```

```
Query Match 1.3%; Score 24; DB 7; Length 172649;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1816 ACTGTAAAAA 1839
Db 126745 ACTGTAAAAA 126768
```

```
RESULT 14
US-11-112-908-65/c
; Sequence 65, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 173115
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-65
```

Query Match 1.3%; Score 24; DB 7; Length 173115;
 Best Local Similarity 100.0%; Pred.No. 0.62;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAAAAAAAAAAAAAAAAA 1839
 Db 18974 ACTGTAAAAAAAAAAAAAAAAAAAA 18951

RESULT 15

US-11-112-908-32
 ; Sequence 32, Application US/11112908
 ; Publication No. US20050260659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Davis, Lisa M.
 ; TITLE OF INVENTION: Breast Cancer Biomarkers
 ; FILE REFERENCE: 04-164-US
 ; CURRENT APPLICATION NUMBER: US/11/112,908
 ; CURRENT FILING DATE: 2005-04-22
 ; PRIOR APPLICATION NUMBER: US 60/564,758
 ; PRIOR FILING DATE: 2004-04-23
 ; PRIOR APPLICATION NUMBER: US 60/575,978
 ; PRIOR FILING DATE: 2004-06-01
 ; PRIOR APPLICATION NUMBER: US 60/631,702
 ; PRIOR FILING DATE: 2004-11-30
 ; PRIOR APPLICATION NUMBER: US 60/633,826
 ; PRIOR FILING DATE: 2004-12-07
 ; NUMBER OF SEQ ID NOS: 511
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 32
 ; LENGTH: 193363
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-11-112-908-32

Query Match 1.3%; Score 24; DB 7; Length 193363;
 Best Local Similarity 100.0%; Pred.No. 0.61;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1816 ACTGTAAAAAAAAAAAAAAAAAAAA 1839
 Db 84066 ACTGTAAAAAAAAAAAAAAAAAAAA 84089

Search completed: December 30, 2005, 09:39:32
 Job time : 302 secs

THIS PAGE BLANK (USPTO)